

Département de Géomatique appliquée
Faculté des Lettres et Sciences Humaines
Université de Sherbrooke

Identification des problèmes phytosanitaires de la vigne au sein de
la parcelle : association de l'imagerie à ultra-haute résolution
spatiale et de l'apprentissage profond

Justine Boulent

Thèse présentée pour l'obtention du grade de Philosophiae Doctor
(Ph. D) en télédétection, Cheminement en géomatique appliquée

18 décembre 2020

© Justine Boulent, 2020

Directeur de recherche :

Pr. Jérôme Théau

Département de Géomatique appliquée, Université de Sherbrooke, Sherbrooke (QC)

Codirecteur de recherche :

Dr. Samuel Foucher

Centre de Recherche en Informatique de Montréal (CRIM), Montréal (QC)

Membres du jury :

Pr. Yacine Bouroubi

Département de Géomatique appliquée, Université de Sherbrooke, Sherbrooke (QC)

Membre interne du jury

Pr. Mickaël Germain

Département de Géomatique appliquée, Université de Sherbrooke, Sherbrooke (QC)

Membre interne du jury

Dr. Nicolas Tremblay

Agriculture et Agroalimentaire Canada, St-Jean-sur-Richelieu (QC)

Membre externe du jury

Résumé

Notre époque est indéniablement marquée par les changements climatiques et la baisse drastique de la biodiversité, questionnant la durabilité de l'ensemble de nos systèmes productifs. En agriculture, la remise en cause des pratiques dites conventionnelles est de plus en plus prégnante. L'usage des pesticides est l'une des pratiques les plus controversées. Bien qu'ayant contribué à l'accroissement sans précédent des rendements agricoles dans les années 1970, ces produits inquiètent désormais par leur dangerosité, autant pour la santé humaine que pour celle de l'environnement. Ils impactent notamment de façon directe la santé des populations d'insectes, dont les pollinisateurs, et du microbiote des sols. Sur le long terme, si l'usage actuel persiste, un déséquilibre des écosystèmes est à craindre. Il y a donc une urgence à sortir du système actuel de gestion de problèmes phytosanitaires. Cette évolution ne va pas être du même niveau de simplicité pour toutes les cultures. La viticulture utilise des quantités importantes de pesticides. En France, en 2006, la vigne représentait 3,3 % de la surface agricole mais 14,4 % des pesticides utilisés. Réaliser des traitements adaptés à la situation phytosanitaire réelle de la parcelle contribuerait à réduire cette consommation. Cependant, connaître l'état de santé précis d'une parcelle donnée n'est pas une information facile à obtenir. Prospector tout un vignoble prend beaucoup de temps, ce qui rend sa réalisation régulière difficile. Dans ce projet, nous souhaitons faciliter la prospection par son automatisation. Celle-ci pourrait se matérialiser par une caméra montée sur un robot, un tracteur ou un drone, dont les images seraient ensuite analysées automatiquement, permettant à l'agriculteur d'obtenir une carte de l'état de santé de ses parcelles. Toutefois, développer un tel outil est loin d'être simple du fait de la complexité des cultures. Celui-ci doit être capable de mener une analyse malgré la diversité des variétés, des stades phénologiques mais aussi des ravageurs, des maladies et de l'ensemble de leurs symptômes. La diversité des conditions d'acquisitions possibles et la complexité du feuillage et de l'arrière-plan constituent également des défis pour l'analyse, empêchant le développement d'un outil de prospection automatique fiable.

Dans ce projet de doctorat, nous avons utilisé de l'apprentissage profond, et plus particulièrement des réseaux de neurones convolutifs, comme technique d'analyse d'images pour la reconnaissance de problèmes phytosanitaires de la vigne. Nous avons plus particulièrement étudié deux maladies : le mildiou et la flavescence dorée. Deux jeux de données conséquents et reflétant la complexité des cultures ont été bâtis à partir de photos Rouge-Vert-Bleu acquises dans des vignobles. L'objectif de ce projet était d'évaluer si les réseaux de neurones convolutifs sont adaptés à l'analyse d'images pour un outil de prospection automatique. Pour cela, tout au long de nos travaux, nous avons analysé la réponse des modèles entraînés à plusieurs

scénarios. Tout d’abord, nous avons confronté les modèles à des images acquises en champ, donc possédant une complexité intrinsèque. Nous avons également évalué la capacité de ces réseaux à reconnaître une maladie possédant différents symptômes présents sur plusieurs organes, le mildiou. Nous avons ensuite évalué la robustesse de l’analyse face au changement de cépages et face au changement des conditions d’acquisition des images, dont un changement de la résolution spatiale et de la plateforme d’acquisition. Nos résultats montrent que les réseaux de neurones convolutifs sont tout à fait appropriés pour la reconnaissance de problèmes phytosanitaires, fournissant des analyses à la fiabilité inédite. Pour la reconnaissance du mildiou et de six autres classes, le modèle développé a obtenu 95,48 % de bonnes prédictions. Quant à celle de la flavescence dorée, le modèle développé a atteint un taux de vrais positifs de 98,48 % sur les images issues de notre propre acquisition, et de 100 % sur un ensemble réduit d’images provenant d’une source externe. Cependant, nous avons également pu identifier plusieurs limites qui restent à surmonter. Principalement, il s’agit de l’incapacité de nos modèles à analyser des images dont le contenu, bien que thématiquement proche, voire similaire, est trop éloigné de celui des images utilisées en entraînement. Néanmoins, plusieurs techniques innovantes, telles que l’apprentissage actif (*active learning*) ou auto-supervisé (*self-supervised learning*), peuvent être adoptées pour surmonter ce problème sans forcément passer par la constitution d’un ensemble d’entraînement annoté parfaitement exhaustif. Ainsi, bien que des améliorations soient nécessaires, le contexte actuel est des plus enthousiasmants pour mener à bien le développement d’un outil de reconnaissance de problèmes phytosanitaires, et par extension, d’outils de prospection automatique pour les vignes, mais aussi pour toutes les autres cultures.

Mots-clés : viticulture et agriculture de précision, reconnaissance automatique de maladies, apprentissage profond, réseaux de neurones convolutifs, mildiou de la vigne, flavescence dorée.

Abstract

Climate change and the severe decline in biodiversity undeniably mark our era, questioning the sustainability of all our production systems. In agriculture, the reconsideration of so-called conventional practices is more and more prevalent. Use of pesticides is one of the most controversial practices. Despite their contribution to the unprecedented increase in agricultural yields in the 1970s, these products are now a matter of concern because of their danger to both human health and the environment. In particular, they directly impact the health of insect populations, including pollinators, as well as soil microbiota. In the long term, if current use persists, imbalance in the ecosystems is to be expected. Therefore, urgent action is needed to move away from the current phytosanitary management system. This change will not be as simple for all crops. Viticulture uses significant quantities of pesticides. In France, in 2006, vineyards covered 3.3 % of the agricultural surface but accounted for 14.4 % of the pesticides used. Applying treatments according to the real health conditions of the field would reduce this consumption. However, it is not simple to know the precise health status of a given plot. Scouting an entire vineyard requires a significant amount of time, which makes it difficult to do so on a regular basis. In this project, we intend to facilitate scouting by its automation. It could be achieved by a camera mounted on a robot, a tractor or a drone, whose images would then be automatically analyzed, providing the farmer with a health map of his fields. However, developing such a tool is not easy given the crops complexity. It must be able to analyze images despite the diversity of varieties, phenological stages, as well as the diversity of pests, diseases and all their symptoms. The variety of acquisition conditions and the complexity of the foliage and background also constitute challenges for the analysis, hindering the development of a reliable automatic scouting tool.

In this study, we used deep learning, more specifically convolutional neural networks, as a technique to analyze images for the recognition of grapevine phytosanitary problems. In particular, we studied two diseases : downy mildew and flavescence dorée. Two large datasets reflecting the complexity of the crops were built from Red-Green-Blue photos taken in vineyards. The goal of this project was to evaluate whether convolutional neural networks are suitable for image analysis for an automatic scouting tool. Therefore, throughout our studies, we analyzed the response of the trained models to several scenarios, firstly to images captured in the field but also to diseases with many symptoms affecting several organs. We also evaluated the robustness of the analysis to changes in grape varieties and to changes in image acquisition conditions, including a change in spatial resolution and acquisition platform. Our results show that convolutional neural networks are well suited for our application, providing unprecedented

reliable analyses. For the recognition of downy mildew and six other classes, the developed model achieved 95.48 % of good predictions. Regarding flavescence dorée, the model developed reached a true positive rate of 98.48 % on images from our own acquisition and 100 % on a reduced set of images from an external source. However, we were also able to identify several limitations that still need to be overcome. Mainly, the inability of our models to analyze images whose content, although thematically close or even similar, is too far away from the images used in training. Nevertheless, several innovative techniques, such as active learning or self-supervised learning, could possibly overcome this problem without the need for a perfectly comprehensive training dataset. Therefore, although improvements are necessary, the current context is most exciting to carry out the development of a phytosanitary problem recognition tool, and by extension, of automatic prospecting tools for grapevines, as well as all other crops.

Keywords : precision viticulture, precision agriculture, automatic plant diseases identification, deep learning, convolutional neural network, downy mildew, flavescence dorée.

Table des matières

Liste des figures	viii
Liste des tableaux	xi
Liste des abréviations	xii
Remerciements	xiv
Avant-propos	xv
1 Introduction	1
1.1 La viticulture dans un contexte de changements environnementaux	2
1.2 Problématique	5
1.3 Objectifs de la recherche	9
1.4 Structure de la thèse	10
2 Cadre théorique : L'apprentissage profond pour la détection des maladies des cultures	14
2.1 Présentation de l'article	15
2.2 Résumé	15
2.3 Convolutional Neural Networks for the Automatic Identification of Plant Diseases : A Review	16
2.3.1 Introduction	16
2.3.2 Research methodology	17
2.3.3 Deep Learning Applied to Diseases Identification	18
2.3.4 Understanding the Trained Models	31
2.3.5 Discussion	33
2.3.6 Conclusion	36
2.3.7 Annex	36
2.3.8 Acknowledgements	39
3 Reconnaissance automatique de maladies de la vigne par classification d'images : Cas d'étude du mildiou	40
3.1 Présentation de l'article	41
3.2 Résumé	42

3.3	Deep learning for in-field image-based grapevine downy mildew identification	42
3.3.1	Introduction	43
3.3.2	Materials and Methods	44
3.3.3	Results and Discussion	48
3.3.4	Conclusion	50
3.3.5	Acknowledgements	51
4	Reconnaissance automatique de maladies de la vigne par segmentation d'images : Cas d'étude de la flavescence dorée	52
4.1	Présentation de l'article	53
4.2	Résumé	55
4.3	Automatic Detection of Flavescence dorée Symptoms across White Grapevines Varieties using Deep Learning	55
4.3.1	Introduction	56
4.3.2	Material and Methods	59
4.3.3	Results	65
4.3.4	Discussion	70
4.3.5	Conclusion	75
4.3.6	Acknowledgments	76
5	Les RNCs face à différentes conditions d'acquisition : Analyse d'images à la résolution spatiale dégradée et acquises par drone	78
5.1	Introduction	79
5.2	Matériel et méthodes	80
5.2.1	Description générale de la méthode	80
5.2.2	Données	82
5.2.3	Méthodes	85
5.3	Résultats	87
5.3.1	Phase d'entraînement	87
5.3.2	Performances des modèles face à une dégradation de résolution spatiale : inférence sur les images acquises par perche	88
5.3.3	Performances des modèles face des images venant de différentes plateformes : inférence sur les images acquises par drone	93
5.4	Discussion et conclusion	98
5.4.1	Les RNCs face au changement de résolution spatiale	98
5.4.2	Les RNCs, d'une plateforme d'acquisition à l'autre	99

6	Discussion générale et conclusion	101
6.1	Contributions à la recherche	102
6.2	Perspectives pour le développement d'outils de prospection automatique des cultures	103
	Références	106
	Annexes	117
	Annexe A : Description des données	118
	Annexe B : Détails techniques	134

Liste des figures

1.1	Illustration de la diversité foliaire d’une maladie : le mildiou de la vigne. (A) Expression dite en tache d’huile. (B) Symptômes avancés : les taches tendent du jaune au brun foncé. (C) Symptômes très avancés : les taches asséchées sont presque totalement brunes foncées. (D) Taches sectionnelles. (E) Taches étendues. (F) Taches suivant les nervures. (G) Mildiou et carences. (H) Mildiou et érinose. (I) Mildiou et embruns de produit phytosanitaire.	6
2.1	Expected output examples of (A) the classification, (B) the object detection, and (C) the segmentation of images containing esca disease symptoms. . . .	19
2.2	Typology of image complexity found in the datasets. Esca grape disease on (A) an image captured under controlled condition (from the PlantVillage dataset), (B) an image captured under uncontrolled condition and with a focus on a particular organ, and (C) an image captured under uncontrolled conditions and without focus on a particular organ.	19
2.3	Selecting an analysis scale : from a scale close up on the main symptoms (A) to a scale providing more contextual features (D). Example of a vine branch affected by <i>flavescence dorée</i>	23
2.4	Comparison of the training time (h) and accuracy values (%) obtained on a validation set according to different architectures and training strategies. Adapted from Brahimi <i>et al.</i> (2018).	26
2.5	Comparison of the accuracies obtained using different architectures. For each study, the lowest and highest performance achieved on the validation set (or test set when available) is reported.	28
2.6	Comparison of accuracy values obtained by CNNs and by other image processing methods. Only the best CNN architecture and comparative approach for each study were reported. If given in the study, the precision on a test set was reported.	30
2.7	Visualization examples found in the corpus (A) Activations in the first convolution layer visualization (Mohanty <i>et al.</i> , 2016) (B) T-distributed Stochastic Neighbour Embedding on the final fully connected layer (Zhang <i>et al.</i> , 2018a)	32
3.1	Illustration of classes and their diversity	46
3.2	Processing setup overview	48
3.3	Original (left) and normalised (right) confusion matrices (%) from ResNet18 evaluation on the test set	50

3.4	Examples of wrong predictions. True Label \Rightarrow Predicted Label. (VLB : Vine Leaf Blister, DM : Downy Mildew, D : Deficiencies, H : Healthy)	50
4.1	Processing setup overview.	60
4.2	Illustration of Flavescence dorée symptoms for the different grape varieties found in the <i>Ext.</i> test dataset. Source : D. Vergnes, FREDON Aquitaine. . .	62
4.3	Examples of samples from the CNN training dataset for FD and Non-FD classes.	63
4.4	Summary of accuracy values obtained on the validation dataset for the 10 runs.	66
4.5	Segmentation maps with minimum and maximum Intersection over Union (IoU) values on Chardonnay datasets. On the <i>Pers. Chardonnay</i> dataset, Images (A) with the lower IoU value, (B) with the higher IoU value. On the <i>Ext. Chardonnay</i> dataset, (C) with the lower IoU value, (D) with the higher IoU value.	68
4.6	Examples of false predictions from the <i>Pers. Chardonnay</i> Test dataset. (A), (B) Undetected Flavescence dorée (FD) symptoms, (C), (D) Over-detection close to FD symptomatic areas, E : Visually unexplainable false FD detection.	69
4.7	Examples of false predictions from <i>Ext.</i> datasets. (A) Original image of Exalta leaf with Flavescence dorée (FD) symptoms ; (B) Prediction associated to image A, only the rolled areas of the leaf are identified as FD ; (C) Undetected FD symptoms on Ugni-Blanc grapevine variety ; (D) Over-detection close to FD symptomatic areas ; (E) False detection on non-grapevine elements ; and (F) Undetected FD symptoms on Exalta grapevine variety.	70
4.8	Segmentation maps with minimum and maximum Intersection over Union (IoU) values on two <i>Ext.</i> datasets. On the <i>Ext. Semillon</i> dataset : Images (A) with the lower IoU value, and (B) with the higher IoU value. On the <i>Ext. Merlot Blanc</i> dataset : (C) with the lower IoU value, and (D) with the higher IoU value.	71
4.9	Illustration of representative features from Guided Grad-CAM that are used by the model to predict the Flavescence dorée (FD) and Non-FD classes. A - C : Non-FD images, D - H : FD images. The predicted label and the source dataset are given next to each image.	73
4.10	Grape varieties in the model space : UMAP visualization of the embeddings based on an arbitrarily-set seed.	74
4.11	Illustration of irrelevant features from GG-CAM that are used by the model to predict the Flavescence dorée (FD) class. A : Trellising wire, petioles, canes, soil and rolled shapes, B : Canes, C : Canes and grapes, D : Trellising wire and triangular shadow shape, E : Trellising wire and triangular shadow shape, and F : Trellising wire and soil.	75

5.1	Illustration des différentes approches d'entraînement et d'inférence adoptées.	82
5.2	Diagramme en boîte représentant la taille des patches en nombre de pixels pour chaque classe, à la fois dans leur version originale et dans leur version réduite par un facteur 3.	83
5.3	Exemples d'images contenues dans l'ensemble de test perche. La première ligne correspond à l'image brute, la seconde à la segmentation manuelle (en clair : les zones présentant de la flavescence dorée).	84
5.4	Exemples d'images contenues dans l'ensemble de test drone. La première ligne correspond à l'image brute, la seconde à la segmentation manuelle (en clair : les zones présentant de la flavescence dorée).	85
5.5	Précisions minimales, maximales et médianes obtenues sur l'ensemble de validation pour les approches 1 et 2, et leurs 5 initialisations.	88
5.6	Exemples de fausses détections de flavescence dorée obtenues lors de l'inférence du meilleur modèle de l'approche 1 sur les images de dimension originale. Les zones sombres sont celles prédites comme "Non flavescence", les zones claires celles prédites comme "Flavescence".	89
5.7	Pourcentages de faux négatifs et de faux positifs obtenus lors de l'inférence sur l'ensemble des images acquises par perche avec les modèles issus de l'approche 1, de l'approche 2 et du Chapitre 4, pour les 3 dimensions étudiées.	92
5.8	Feuilles présentant des symptômes de flavescence dorée n'ayant pas été détectés lors de l'inférence sur une dimension similaire à celle d'images drone avec les trois approches (A, B, C), et avec l'approche 1 et celle du Chapitre 4 (D). . .	93
5.9	Pourcentages de faux négatifs et de faux positifs lors de l'inférence sur les images acquises par drone avec les modèles issus de l'approche 1, de l'approche 2 et du Chapitre 4.	96
5.10	Exemples de zones de flavescence dorée non repérées par le modèle de l'approche 2 sur les images acquises par drone.	96
5.11	Inférence des modèles issus des deux approches étudiées et du cas d'étude développé dans le Chapitre 4 sur une image drone, en dimensions originale et équivalente à un vol à 25 mètres d'altitude.	97
6.1	Résultat de l'alignement des canaux par calcul d'une fonction homographique sur une image acquise au moyen de la Micasense Red-Edge.	119
6.2	Perche en cours d'utilisation.	120
6.3	Exemples de photos prises avec la perche.	120
6.4	Exemples de photos prises avec la perche.	121
6.5	Exemples de photos prises avec le drone.	121

6.6	Localisation des régions d'acquisition en France et au Canada.	122
6.7	Localisation des deux parcelles plantées en vignes (G18 et G22) à la ferme expérimentale d'Agriculture et Agroalimentaire Canada (AAC) à Frelighsburg, d'après une carte fournie par AAC.	123
6.8	Expression du mildiou sur inflorescences et tiges. (A) Rameau sain (gauche) et rameau contaminé par le mildiou, stade d'infection primaire (droite). (B) Inflorescence contaminée par le mildiou, stade des sporulations. (C) Inflorescence contaminée par le mildiou, stade avancé des sporulations.	126
6.9	Évolution des symptômes foliaires de mildiou : (A) Taches d'huile. (B) Symptômes avancés : les taches tendent du jaune foncé au brun. (C) Symptômes très avancés : les taches sont presque complètement brunes.	126
6.10	Diversité des expressions du mildiou : (A) Taches sectionnelles. (B) Taches étendues. (C) Taches suivant les nervures.	127
6.11	Association du mildiou avec d'autres problèmes phytosanitaires : (A) Mildiou et carence. (B) Mildiou et l'érinose. (C) Mildiou et embruns de produits de traitement.	127
6.12	Flavescence dorée sur cépages blancs : (A),(B) Chardonnay. (C) Chenin. (D) Ugni-Blanc.	128
6.13	Flavescence dorée sur cépages rouges : (A) Grenache. (B) Syrah. (C) Cabernet. (D) Non identifié.	128
6.14	Association de symptômes de flavescence dorée avec un autre problème phytosanitaire : (A) Flavescence et mildiou. (B) Flavescence et carence.	129
6.15	Dégâts causés par des cicadelles : (A) Espèce non connue, (B) Cicadelle bubale	129
6.16	Carences : (A) En magnésium sur cépage blanc. (B) En magnésium sur cépage rouge. (C) En manganèse. (D) En potassium sur cépage rouge. (E) En potassium sur cépage blanc. (F) En fer (peu avancé).	130
6.17	Esca sur : (A) Cépage rouge. (B) Cépage rouge. (C) Cépage blanc.	130
6.18	Black Dead Arm : (A) (B) (C) Trois expressions foliaires différentes. (D) Forme apoplectique de BDA ou d'esca.	130
6.19	Dégâts provoqués par des ravageurs : (A) Erinose. (B) Marques laissées par des acariens. (C) Marques laissées par des scarabées japonais.	131
6.20	Marques laissées par les traitements : (A) Phytotoxicité. (B) et (C) Embruns de traitements.	131
6.21	Exemples de problèmes phytosanitaires non identifiés.	132
6.22	Exemples de différentes conditions d'illumination rencontrées.	133

6.23	Différents angles d'acquisition et distance au feuillage : (A) Visée nadir, acquisition éloignée. (B) Visée nadir, acquisition rapprochée. (C) Visée latérale, acquisition éloignée. (D) Visée latérale, acquisition rapprochée.	133
------	--	-----

Liste des tableaux

1.1	Descriptif des trois cas d'étude présentés dans cette thèse.	13
2.1	Summary table – Part 1	37
2.2	Summary table – Part 2	38
3.1	Samples origin and grape varieties per dataset and class (in %). The percentage distribution between grape varieties in Frelighsburg is broadly estimated. . .	45
3.2	Samples per dataset and class.	46
3.3	Accuracy of the best models for each ResNet (%).	49
3.4	Classification report from ResNet-18 evaluation on the test set.	49
4.1	Summary of datasets used to train the CNN classifier and to evaluate the FCN model. *As we extracted samples for both classes in some images, we only indicate the total images' count for the CNN Training dataset.	63
4.2	Hyperparameters used to trained the ResNet-18 classifier.	64
4.3	Summary of inference results on both Chardonnay datasets, presenting the True Positive Rate (TPR), False Positive Rate (FPR), mean, min and max values of Intersection over Union (IoU). Acronyms : TP : True Positive, P : Positives, FP : False Positive, N : Negatives. * IoU min, considering only the True Positive predictions.	67
4.4	Summary of inference results on other grapevine varieties' datasets, presenting the True Positive Rate (TPR), False Positive Rate (FPR), mean, min and max values of Intersection over Union (IoU). Acronyms : TP : True Positive, P : Positives, FP : False Positive, and N : Negatives. * IoU min, considering only the True Positive predictions.	69
5.1	Hyperparamètres utilisés pour entraîner les modèles basés sur l'architecture ResNet-18.	86
5.2	Synthèse des résultats d'inférence obtenus avec les modèles entraînés avec l'approche 1 sur les trois dimensions d'image étudiées. Des cinq initialisations d'entraînement réalisées, seuls les modèles obtenant la valeur minimale, médiane et maximale de F1-Score sont présentés. Les nombres en gras correspondent aux valeurs de précision, rappel ou F1-score les plus élevées. VP = Vrai Positif, P = Positif, FP = Faux Positif, N = Négatif, mIoU = <i>mean Intersection over Union</i>	89

5.3	Synthèse des résultats d'inférence obtenus avec la seconde approche sur les trois dimensions d'image étudiées. Des cinq initialisations d'entraînement réalisées, seuls les modèles obtenant la valeur minimale, médiane et maximale de F1-Score sont présentés. Les nombres en gras correspondent aux valeurs de précision, rappel ou F1-score les plus élevées. VP = Vrai Positif, P = Positif, FP = Faux Positif, N = Négatif, mIoU = <i>mean Intersection over Union</i>	90
5.4	Comparaison des résultats d'inférence obtenus avec le modèle au F1-score le plus élevé pour les deux approches ainsi que le modèle développé dans le Chapitre 4. Les nombres en gras correspondent aux valeurs de précision, rappel ou F1-score les plus élevées. VP = Vrai Positif, P = Positif, FP = Faux Positif, N = Négatif, mIoU = <i>mean Intersection over Union</i>	91
5.5	Synthèse des résultats d'inférence obtenus avec la première approche sur les deux dimensions d'images acquises par drone étudiées. Des cinq initialisations d'entraînement réalisées, seuls les modèles obtenant la valeur minimale, médiane et maximale de F1-Score sont présentés. Les nombres en gras correspondent aux valeurs de précision, rappel ou F1-score les plus élevées. VP = Vrai Positif, P = Positif, FP = Faux Positif, N = Négatif, mIoU = <i>mean Intersection over Union</i>	94
5.6	Synthèse des résultats d'inférence obtenus avec la seconde approche sur les deux dimensions d'image acquises par drone étudiées. Des cinq initialisations d'entraînement réalisées, seuls les modèles obtenant la valeur minimale, médiane et maximale de F1-Score sont présentés. Les nombres en gras correspondent aux valeurs de précision, rappel ou F1-score les plus élevées. VP = Vrai Positif, P = Positif, FP = Faux Positif, N = Négatif, mIoU = <i>mean Intersection over Union</i>	94
5.7	Comparaison des résultats d'inférence obtenus avec le modèle au F1-score le plus élevé pour les deux approches ainsi que le modèle développé dans le Chapitre 4 sur les images acquises par drone. Les nombres en gras correspondent aux valeurs de précision, rappel ou F1-score les plus élevées. VP = Vrai Positif, P = Positif, FP = Faux Positif, N = Négatif, mIoU = <i>mean Intersection over Union</i>	95
6.1	Nombre d'images selon leur localisation et par type d'acquisition.	125
6.2	Nombre d'images selon la couleur de cépage et par type d'acquisition. Les photos prises en Montérégie en 2017 n'ont pas pu être incluses à ce tableau de synthèse.	125
6.3	Caractéristiques des ordinateurs utilisés pour entraîner et évaluer les modèles.	134

6.4	Principales librairies utilisées dans le développement des algorithmes	134
6.5	Temps d'entraînement des ResNet pour le cas d'étude sur la reconnaissance du mildiou	135
6.6	Temps d'entraînement des ResNet-18 selon le jeux d'entraînement utilisé dans le cas d'étude évaluant les performances des RNCs face à un changement de résolution spatiale.	135
6.7	Temps d'inférence par image selon la dimension des images dans le cas d'étude évaluant les performances des RNCs face à un changement de résolution spatiale. Les inférences ont été réalisées sur le PC1.	135

Liste des abréviations

AP : Apprentissage Profond.

B : *Background*.

D : *Deficiencies*.

DM : *Downy Mildew*.

DPO : *Dry Plant Organ*.

FAO : Organisation des Nations unies pour pour l'alimentation et l'agriculture (*Food and Agriculture Organization of the United Nations*).

FD : Flavescence dorée.

FN : Faux Négatif (*False Negative*).

FP : Faux Positif (*False Positive*).

FPR : *False Positive Rate*.

FREDON : Fédération Régionale de Défense contre les Organismes Nuisibles.

GG-CAM : *Guided Gradient-weighted Class Activation Mapping*.

H : *Healthy*.

IA : Intelligence Artificielle.

IoU : *Intersection over Union*.

JB : *Japanese Beetle*.

NIR : *Near Infrared*.

NLB : *Northern Leaf Blight*.

RNC : Réseau de Neurones Convolutifs (*CNN : Convolutional Neural Network*).

RPC : Réseau Pleinement Convolutif (*FCN : Fully Convolutional Network*).

RVB : Rouge-Vert-Bleu (*RGB : Red-Green-Blue*).

TPR : *True Positive Rate*.

t-SNE : *t-distributed Stochastic Neighbour Embedding*.

UNESCO : Organisation des Nations unies pour l'éducation, la science et la culture (*United Nations Educational, Scientific and Cultural Organization*).

UMAP : *Uniform Manifold Approximation and Projection*.

VLB : *Vine Leaf Blister*.

VP : Vrai Positif (*TP : True Positive*).

Remerciements

Cette dernière année passée à travailler depuis chez soi fait d'autant plus ressortir la valeur de son entourage, notamment pour mener à bien les projets dans lesquels nous nous engageons. Ce projet de doctorat arrive aujourd'hui à son terme, notamment grâce à tous les organismes et personnes qui suivent.

Je remercie Jérôme Théau, Professeur au Département de Géomatique appliquée de l'Université de Sherbrooke, pour la confiance qu'il m'a accordée en me permettant d'effectuer mes travaux de recherche sous sa direction. Son suivi régulier de mes avancées m'a assurément aidée à garder le cap pour mener ces travaux à aboutissement.

Je remercie Samuel Foucher, Chercheur au Centre de Recherche en Informatique de Montréal, qui a accepté de co-diriger ce projet. Cette collaboration a été un apport déterminant pour la réalisation de mes travaux. Ses conseils avisés et sa grande implication dans le projet ont été très précieux et motivants.

Mes remerciements vont également aux organismes qui ont financé ce projet : le Centre de Géomatique du Québec, Mitacs, la compagnie Microdrones et l'Université de Sherbrooke.

Je remercie également le CRIM pour son accueil, et plus particulièrement l'équipe VISI des premières heures – une gang full fine qui a réussi à rendre mes journées (semaines ? mois ?) d'annotation moins plates et grâce à qui je maîtrise à 100 % le joual d'icitte. Un remerciement particulier à Mario Beaulieu et à Pierre-Luc St-Charles qui m'ont guidée dans mes premiers pas en apprentissage automatique. Le chemin aurait été bien plus ardu sans vous !

Je remercie également toutes les personnes qui ont contribué à mes campagnes d'acquisition photographiques, aussi bien ceux qui ont accepté que j'accède à leurs cultures, que ceux qui ont pris le temps de partager leur savoir en me formant à la reconnaissance des problèmes phytosanitaires. Je remercie également la compagnie Microdrones qui a mis tous les moyens à ma disposition pour que ces acquisitions conduisent aux images désirées.

Sur une note plus personnelle, je tiens également à remercier mes amis, ceux qui partagent mon quotidien au Québec (avec une mention spéciale à ceux du Jardin Botanique, qui partagent aussi les aléas de la recherche), mais aussi ceux qui sont en France, et que j'ai tellement hâte de retrouver à chaque vacances. Un grand merci à ma famille, et notamment aux petits Léon et Tom, qui même à distance m'apportent beaucoup de joie et de rires. Être loin de vous tous a sûrement été l'une des choses les plus difficiles dans ce doctorat.

Pour finir, le plus grand des mercis à Simon, pour m'avoir accompagnée dans cette aventure canadienne (qui risque de se poursuivre au moins encore un peu).

Avant-propos

Ce manuscrit est présenté sous la forme d'une thèse par articles et se compose de six chapitres, dont trois sous forme d'article (Chapitres 2, 3 et 4). Ces articles sont présentés dans leur langue de publication, en anglais, et sont précédés d'une description introductive en français. Le Chapitre 1 constitue le chapitre introductif et présente le contexte, la problématique et les objectifs de ce projet de doctorat.

Dans le Chapitre 2, une revue de littérature sur l'utilisation des réseaux de neurones convolutifs pour l'identification automatique des maladies des cultures est présentée. Cette revue de littérature a été publiée dans le journal *Frontiers in Plant Science*, en juillet 2019 : Boulent, J., Foucher, S., Théau, J., St-Charles, P.-L. (2019) Convolutional Neural Networks for the Automatic Identification of Plant Diseases : A Review. *Frontiers in Plant Science*, Technical Advances in Plant Science, 10, 941, doi :10.3389/fpls.2019.00941.

Le Chapitre 3 présente un cas d'étude sur la reconnaissance des symptômes de mildiou par classification d'images. Il a été présenté lors de la *12th European Conference on Precision Agriculture*, qui s'est tenue à Montpellier en juillet 2019, puis a été publié dans les actes de congrès : Boulent, J., Beaulieu, M., St-Charles, P.-L., Théau, J., Foucher, S., (2019) Deep learning for in-field image-based grapevine downy mildew identification. *Precision agriculture '19. Proceedings of the 12th European Conference on Precision Agriculture*, Montpellier, France, Wageningen Academic Publishers, pp. 141-148, doi : 10.3920/978-90-8686-888-9_16.

Le Chapitre 4 présente un cas d'étude sur la reconnaissance des symptômes de flavescence dorée par segmentation d'images. Cet article a été publié dans le journal *Frontiers in Artificial Intelligence*, en novembre 2020 : Boulent, J., St-Charles, P.-L., Foucher, S., Théau, J. (2020) Automatic detection of Flavescence dorée symptoms across white grapevine varieties using deep learning, doi : 10.3389/frai.2020.564878.

Le Chapitre 5 présente une étude sur l'impact de deux paramètres d'acquisition d'images (la résolution spatiale et la plateforme utilisée) sur la performance des réseaux de neurones convolutifs.

Enfin, le Chapitre 6 présente une discussion générale et une conclusion, ouvrant également sur des perspectives de recherche pour la reconnaissance automatique des problèmes phytosanitaires.

Chapitre 1

Introduction

1.1 La viticulture dans un contexte de changements environnementaux

La viticulture constitue un secteur agricole dynamique dans plusieurs pays à travers le monde. Initialement cantonnée aux territoires européens et moyen-orientaux, elle s'est étendue au-delà de ces milieux de production traditionnels au cours du XX^{ème} siècle, majoritairement au sein de pays de l'hémisphère sud, comme en Australie, en Argentine, au Chili ou en Afrique du Sud, mais également au sein de l'hémisphère nord comme en Chine, aux États-Unis ainsi qu'au Canada (Johnson et Robinson, 2013; Reynolds, 2017; Outreville, 2009). Cette extension géographique a contribué à faire évoluer le secteur viticole : nouvelles techniques de viticulture, de vinification, nouvelles associations de cépages et nouvelles identités viti-vinicoles. En 2018, la production de vin mondiale s'élevait à 292 millions d'hectolitres et le commerce international du vin, soit la somme des exportations de tous les pays, pesait pour 31,4 milliards d'euros (Blasco, 2019). Preuve du dynamisme du secteur viticole, ces chiffres ne suffisent cependant pas à rendre compte de l'importance singulière de la vigne au sein des territoires qui la cultivent. La vigne, particulièrement dans les lieux d'implantation traditionnels, est très liée à l'histoire et à la culture locales, rythmant en partie la vie de la communauté avec les nombreux événements liés à la production viticole (La Cité du Vin, 2020). La vigne est également une culture particulière du fait de son importance dans le façonnement des paysages. A travers le monde, on trouve une immense variété de paysages viticoles, dont la diversité est liée à la manière dont les vigneron ont adapté la culture de la vigne en fonction du type de sol, de l'exposition, du relief ou encore des cépages (Luginbühl, 2005). Certains paysages viticoles ont été reconnus comme faisant partie du patrimoine mondial de l'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO), comme le paysage de l'île de Pico, au Portugal, celui du Piémont, en Italie ou encore le vignoble en terrasse de Lavaux, en Suisse (UNESCO, 2020). Cette singularité historique, culturelle et paysagère a servi de socle au développement de l'œnotourisme, une activité en pleine croissance. En France, première destination touristique internationale, ce sont 10 millions de personnes qui ont réalisé au moins une activité œnotouristique durant leur séjour en 2016, soit 33 % de plus qu'en 2009 (Atout France, 2020). Les dépenses globales de ces touristes ont été estimées à 5,2 milliards d'euros (Direction Générale des Entreprises, 2017), contribuant ainsi au dynamisme du secteur viticole pour toute l'économie des régions productrices.

Cependant, la viticulture est aujourd'hui confrontée à plusieurs défis liés aux changements environnementaux, risquant d'impacter le dynamisme actuel du secteur. Avec la hausse des températures, celle du CO₂ atmosphérique et avec la raréfaction de l'eau, des impacts directs sur les conditions de culture et d'élevage sont à prévoir (Hinneewinkel, 2010). Dans tous

les scénarios de hausse de températures désormais envisageables, les conditions de terroirs seront modifiées. Deux évolutions directes sont notamment attendues : (1) une avancée des stades phénologiques, (2) un impact sur la composition chimique des baies et donc sur leur profil aromatique (Seguin, 2007; Darriet *et al.*, 2016; Van Leeuwen et Darriet, 2016). Dans l'hémisphère nord, il est estimé qu'une hausse des températures d'un degré entraînerait un déplacement des régions viticoles de 180 kilomètres vers le nord. Cela permettrait à de nouvelles régions de cultiver la vigne, mais en parallèle, les vignobles méridionaux actuels risqueraient de disparaître en grande partie (Hinnewinkel, 2010). La carte de production mondiale des vins s'en trouverait ainsi bouleversée, remettant en question l'ancrage territorial et les identités viticoles actuels.

Cette évolution climatique devrait également bouleverser la répartition géographique de certains ravageurs et maladies. Avec l'augmentation des températures, des insectes et agents pathogènes pourraient en effet s'établir dans de nouveaux environnements mais aussi mieux survivre à l'hiver. Pour les insectes, dont le développement est grandement influencé par la température, une croissance plus rapide et un plus grand nombre de générations par année sont également attendus. Une désynchronisation entre le développement d'insectes et de leurs prédateurs pourrait également survenir, augmentant le risque de dégâts dans les parcelles si cela concerne des insectes ravageurs (Pimentel et Greiner, 1997; Brodeur *et al.*, 2013; Saguez, 2017). L'ensemble des conséquences de l'évolution climatique sur les insectes et agents pathogènes est difficile à prévoir car celles-ci résultent de mécanismes complexes. Cependant, on peut s'attendre globalement à une pression phytosanitaire exacerbée, ou du moins différente (Rosenzweig *et al.*, 2001; Brodeur *et al.*, 2013). Selon l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO), les ravageurs et maladies entraînent de 20 % à 40 % de pertes de récolte, constituant déjà un risque pour la sécurité alimentaire (FAO, 2017). Avec les changements climatiques, certaines régions pourraient voir leurs rendements agricoles décroître, et l'augmentation de la pression phytosanitaire risque d'exacerber ces réductions (Rosenzweig *et al.*, 2001). Il y a donc une nécessité à adapter, réinventer les pratiques de protection actuelles afin de limiter la vulnérabilité aux ravageurs et maladies des cultures de demain.

En parallèle, les exigences environnementales deviennent plus élevées envers le secteur agricole et remettent d'autant plus en question les pratiques actuelles de protection des cultures. L'usage systématique et uniforme de pesticides présente en effet de lourdes conséquences sur la biodiversité, autant en surface que dans le sol (Wilson et Tisdell, 2001). Il entraîne aussi des pollutions de l'air et des cours d'eau (Risebrough, 1986; Gill et Garg, 2014; Goulson, 2014; Sanchez-Bayo et Goka, 2014; Knillmann et Liess, 2019), et est également néfaste pour la santé humaine (Weisenburger, 1993; Bassil *et al.*, 2007; Kim *et al.*, 2016). Au long terme, à

cause de cet usage des pesticides, jumelé aux autres pratiques caractéristiques des systèmes dits conventionnels (comme la monoculture ou le labour), l'équilibre des agrosystèmes risque d'être affecté (Kumar *et al.*, 2020). Ainsi, en cas de surexploitation ou de contamination de l'environnement, on peut s'attendre à une baisse de la qualité et de la quantité des récoltes. De plus, ce système coûte cher à l'agriculteur. A titre d'exemple, selon le Référentiel économique du vigneron 2015-2019 (Chambre d'Agriculture de Loir-et-Cher, 2015), pour un vignoble en pratique conventionnelle de 20 hectares, produisant 55 hectolitres à l'hectare, la protection phytosanitaire est estimée à 486 euros par hectare, soit 30 % du total des coûts de production. Le système conventionnel, n'est durable ni pour l'environnement, ni pour l'économie agricole, ni pour la santé humaine. Il est donc nécessaire de faire évoluer ces pratiques agricoles. Certains gouvernements – comme celui de la France (Ministère de l'Écologie et du Développement durable et Ministère de l'Agriculture et de l'Alimentation, 2015) ou du Québec (Ministère de l'Environnement, du Développement durable et de la Lutte contre les changements climatiques, 2015) – officialisent ces exigences de durabilité au travers de lois et de plans d'action. Concrètement, ces textes visent par exemple à limiter les zones et périodes d'épandage, à interdire l'utilisation de certaines molécules ou encore à réduire les quantités de pesticides utilisées. La viticulture consomme beaucoup d'intrants et se trouve donc particulièrement concernée. En France, en 2006, la vigne représentait 3,3 % de la surface agricole mais 14,4 % des pesticides utilisés (Butault *et al.*, 2011). Pour environ 80 %, ces pesticides utilisés en viticulture sont des fongicides (Ambiaud, 2012). En viticulture, la lutte conventionnelle se fie à un calendrier de traitements réguliers. Les épandages se font généralement de manière systématique, sans forcément prendre en compte ni la météo ni la pression phytosanitaire réellement présente sur la parcelle (Carisse, 2009).

Répondre aux exigences de durabilité peut cependant se révéler complexe pour un agriculteur, surtout s'il vise le maintien immédiat des niveaux de qualité et de quantité de production actuels. Financièrement, sortir de ce système n'est pas simple, surtout quand le prix de la récolte n'est pas ou trop peu revalorisé (Wilson et Tisdell, 2001). Les objectifs de réduction de pesticides sont encore difficilement tenus dans certains pays. En France, le nombre de doses unités triennal (NODU – un indicateur utilisé pour évaluer la consommation de produits phytosanitaires) a augmenté de 12,4 % entre 2009-2010-2011 et 2014-2015-2016, à l'opposé des objectifs fixés par le plan Ecophyto II (Ministère de la Transition Ecologique et Solidaire et Ministère de l'Agriculture et de l'Alimentation, 2018). Pourtant, des stratégies phytosanitaires alternatives font leurs preuves et permettent une production agricole durable. L'une d'elle est celle de la protection intégrée des cultures. Elle correspond à un ensemble de méthodes de lutte biologique, physique et chimique, qui sont en grande partie préventives et qui visent à empêcher le développement de ravageurs et maladies. L'utilisation de pesticides n'est pas

prohibée mais est seulement utilisée de manière réactive et rationnelle (Lucas, 2007). Pour la vigne, la mise en place de cette stratégie de protection commence avant même la plantation du vignoble, avec le choix du site, de l'orientation, des opérations de drainage ou encore du choix de cépages. Des mesures sont également à prendre tout au long de la saison pour limiter la vulnérabilité de la vigne, comme l'élimination de souches mortes, l'effeuillage régulier ou encore le traitement des nouveaux plants à l'eau chaude (Coulon et Hugueniot, 2008). Les opérations au sein de la parcelle sont décidées selon la réalité phytosanitaire de celle-ci, qui est connue grâce aux prospections. Ces inspections des parcelles sont l'élément clef pour détecter le plus rapidement possible les signes de problèmes phytosanitaires, permettant d'agir en conséquence pour protéger les cultures (Carisse, 2009). Cependant, les prospections ne sont pas réalisées de façon rigoureuse dans tous les vignobles. Elles peuvent être incomplètes, pas assez régulières voire inexistantes. La principale cause à cette absence d'adoption généralisée est la contrainte que les prospections représentent. En effet, prospecter l'ensemble d'un vignoble prend beaucoup de temps et devient très difficile à gérer pour de grandes superficies. Pourtant, une prospection régulière est conseillée afin de pouvoir suivre au mieux le vignoble et limiter ainsi sa vulnérabilité – au moins une fois par semaine (Barriault, 2012) et, idéalement, deux fois du débourrement à la récolte (Carisse *et al.*, 2006). Réfléchir à un moyen de faciliter la réalisation des prospections est donc un axe pertinent pour aider à une transition vers des pratiques agricoles plus durables. En effet, avec une adoption régulière des prospections, les agriculteurs possèdent une vue d'ensemble précise de la situation phytosanitaire de leurs parcelles. Ainsi, leurs pratiques peuvent être adaptées aux besoins réels de leurs cultures, permettant de les protéger sans forcément passer par les épandages systématiques préventifs.

1.2 Problématique

La viticulture se trouve dans une période charnière. Pour les professionnels de ce secteur qui risque d'être bouleversé par les changements environnementaux, il s'agit (1) de se prémunir face à l'augmentation de l'exposition de leurs parcelles aux problèmes phytosanitaires, (2) de transformer leur système de protection des cultures pour le rendre moins gourmand en pesticides, et (3) d'obtenir les niveaux de production et de qualité nécessaires à leur activité économique. Pour maîtriser un risque, sa connaissance et son suivi sont primordiaux (Cardona, 2003) – d'où l'importance des prospections. Obtenir ces observations de façon régulière suppose cependant des opérations assez conséquentes dans la gestion quotidienne des vignobles (Coulon et Hugueniot, 2008). De surcroît, les prospections ne sont pas toujours très efficaces. Ce sont des opérations qui prennent beaucoup de temps mais qui ne permettent pourtant pas une surveillance intégrale des parcelles – à défaut de faire toute la parcelle, la stratégie

adoptée peut être la prospection d'un rang sur deux, voire de réaliser uniquement le tour de la parcelle. De plus, les dépisteurs ne sont pas toujours suffisamment formés pour cette tâche (Coulon et Hugueniot, 2008). Identifier les problèmes phytosanitaires rencontrés dans une parcelle est loin d'être une tâche facile. En effet, les symptômes d'un problème donné vont évoluer dans le temps et vont également avoir des apparences différentes selon leur intensité. Plusieurs problèmes phytosanitaires sont également susceptibles d'apparaître sur la même plante, rendant leur identification encore plus complexe (Figure 1.1) (Carisse *et al.*, 2006; Barbedo, 2016). Une bonne connaissance de la parcelle, de ses sensibilités et de l'historique de contamination peut donc aider l'agriculteur lors du diagnostic (Carisse, 2009). Pour acquérir cette expertise, des prospections régulières et sur plusieurs saisons sont nécessaires, ce qui peut être coûteux en terme de formation.

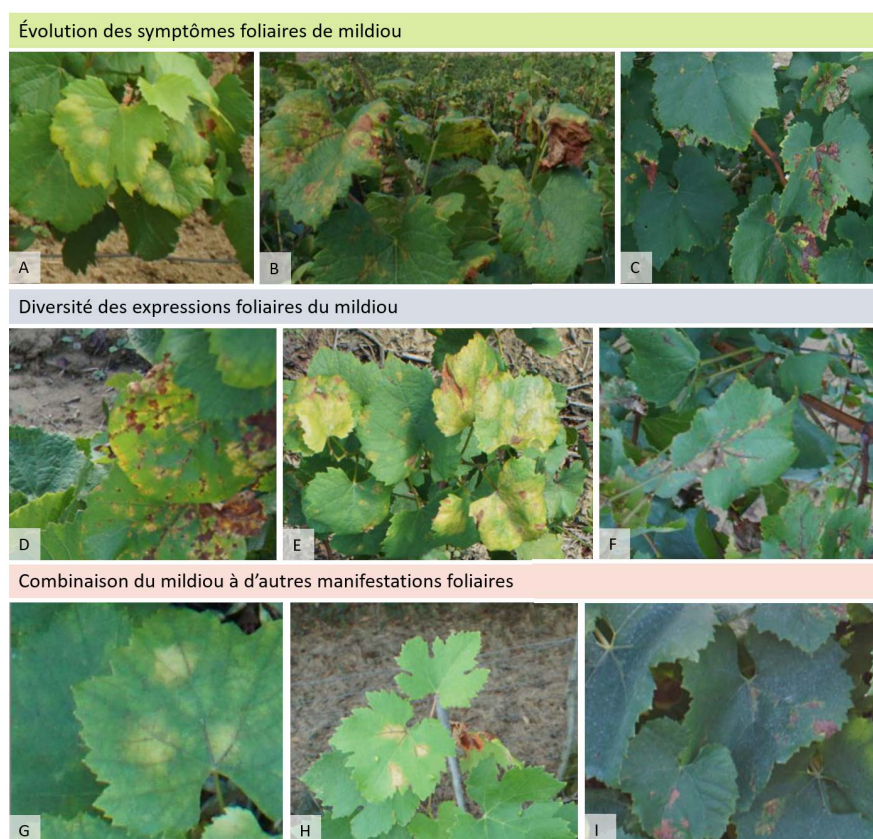


FIGURE 1.1 – Illustration de la diversité foliaire d'une maladie : le mildiou de la vigne. (A) Expression dite en tache d'huile. (B) Symptômes avancés : les taches tendent du jaune au brun foncé. (C) Symptômes très avancés : les taches asséchées sont presque totalement brunes foncées. (D) Taches sectionnelles. (E) Taches étendues. (F) Taches suivant les nervures. (G) Mildiou et carences. (H) Mildiou et érinose. (I) Mildiou et embruns de produit phytosanitaire.

L'automatisation des prospections pourrait être une solution pour rendre cette tâche moins pénible et énergivore, et donc plus adoptée et systématique, mais aussi pour rendre la

prospection plus complète et plus fiable. En effet, en développant un modèle grâce à l’expertise de spécialistes, on peut viser à ce que l’outil de prospection devienne aussi précis qu’eux dans leurs diagnostics. En médecine, plusieurs outils d’analyse d’images pour la réalisation de diagnostics ont été développés ces dernières années, permettant d’obtenir des résultats comparables aux analyses effectuées par des spécialistes (Litjens *et al.*, 2017; Rong *et al.*, 2020). Surtout, ces outils permettent aux spécialistes de gagner du temps, soit en automatisant des diagnostics évidents ou en guidant leur analyse. Un outil similaire pourrait être imaginé pour les prospections dans les vignobles : grâce à une plateforme munie d’un capteur d’images, les vignes présentant des symptômes pourraient être détectées et un diagnostic cartographique proposé. Avec ces informations spatialisées, l’agriculteur pourrait alors choisir la mise en place d’interventions physiques, biologiques ou chimiques afin de protéger ses vignes de manière optimale. Un tel outil viendrait ainsi enrichir, améliorer, faciliter les prospections – en somme, deviendrait les nouveaux yeux des exploitants.

L’idée d’utiliser des images spatialisées pour identifier des problèmes phytosanitaires est loin d’être novatrice. La télédétection est utilisée depuis des années pour développer des applications de diagnostic ou d’identification de zones agricoles endommagées par des infestations (Curran, 1985; Brisco *et al.*, 1998; Hall *et al.*, 2002; Mulla, 2013). Les améliorations techniques récentes en font une approche d’autant plus pertinente pour intégrer la routine de protection phytosanitaire des cultures. Ces vingt dernières années, la résolution spatiale des capteurs a fortement augmenté alors que le coût d’acquisition de la donnée n’a cessé de diminuer (Martinelli *et al.*, 2015). L’autonomie de l’acquisition d’images s’est également accrue grâce au déploiement de plateformes telles que les drones, les robots ou encore les tracteurs avec capteurs embarqués. Ces plateformes permettent d’acquérir rapidement des images à très haute résolution spatiale, voire proximale, sur l’ensemble de la parcelle. Cela représente donc une configuration idéale pour le suivi de la situation phytosanitaire des cultures, d’autant plus lorsque la plateforme réalise son acquisition de façon autonome ou lors des opérations agricoles de routine de l’agriculteur, lui demandant ainsi peu d’efforts supplémentaires pour son déploiement. Même si les niveaux d’adoption sont très inégaux d’un pays et d’une culture agricole à l’autre, certaines applications ont déjà été adoptées par des agriculteurs, notamment la délimitation de zones de gestion intra-parcellaires en se basant sur des cartes de rendement ou de qualité des sols (Zhang *et al.*, 2002; Schimmelpfennig, 2016). Les technologies géospatiales sont donc familières au milieu agricole.

Généralement, pour les applications agricoles, les capteurs utilisés acquièrent des images dans des domaines spectraux allant du visible au proche-infrarouge. Du fait de la signature spectrale des végétaux, des informations sur l’état de santé et de développement des plantes sont obtenues (Knipling, 1970; Kumar *et al.*, 2001). Selon Martinelli *et al.* (2015), quatre

catégories de techniques sont majoritairement utilisées pour analyser la réflectance des végétaux dans le but de détecter les maladies des végétaux : (1) des analyses de corrélation et de régression entre la présence ou la sévérité d'une maladie donnée et la réponse spectrale de la plante dans des longueurs d'ondes spécifiques, (2) la définition d'indices de végétation généraux ou spécifiques, (3) l'utilisation de techniques d'exploration des données pour la réduction de dimension des données spectrales afin de ne sélectionner que les caractéristiques les plus pertinentes et (4) l'utilisation de techniques d'apprentissage machine de toutes sortes (paramétriques ou non, supervisées ou non) pour la réalisation de classification pixel à pixel. Une autre approche mise en avant est l'inversion de modèle de transfert radiatif afin d'extraire des caractéristiques biochimiques, telle que la concentration de chlorophylle, et structurales, comme l'indice de surface foliaire, dont les valeurs peuvent être impactées en cas de stress végétal.

Cependant, malgré l'ensemble des travaux menés, peu d'outils de prospection automatique sont disponibles en agriculture. L'une des raisons de cette faible disponibilité est la complexité du développement de tels outils. En effet, toute la difficulté de prodiguer un diagnostic s'intensifie lorsqu'il doit se baser uniquement sur une image ou un spectre de réflectance. Tous les défis liés au diagnostic cités précédemment, comme l'évolution de l'expression des symptômes, leurs possibles combinaisons ou encore la proximité d'expression de certaines maladies, s'en trouvent exacerbés. La différenciation des stress subis par les végétaux se pose comme un véritable problème, devant pourtant être surmonté pour assurer la fiabilité de la prospection. Ainsi, la vaste majorité des études se concentre sur la détection d'une unique maladie parmi un ensemble restreint de problèmes phytosanitaires, ce qui n'est pas adapté à une prospection automatique complète et opérationnelle (Barbedo, 2016; Sankaran *et al.*, 2010). A cela s'ajoute d'autres défis liés à l'analyse d'images acquises en champ, comme la gestion d'arrière-plans complexes ou de conditions d'acquisition variables (Barbedo, 2016). Ainsi, la télédétection permet certes d'obtenir de l'information spatialisée sur l'état de santé des plantes d'une parcelle mais ne permet pas encore de formuler des diagnostics précis et fiables. Parvenir à mettre en place un outil de prospection automatique paraît difficile avec les approches présentées jusqu'alors.

En parallèle, depuis les années 2010, l'intelligence artificielle (IA) conduit à une véritable révolution technologique, notamment en reconnaissance d'images. L'IA est un terme générique qui réfère à l'ensemble des théories et techniques visant au développement de systèmes informatiques ayant des capacités d'analyse, d'actions rationnelles, similaires à l'intelligence humaine (Russell et Norvig, 1995). L'IA existe depuis les années 1950 mais c'est du fait des avancées en apprentissage machine, avec particulièrement l'obtention de nouvelles capacités de calcul et le développement de l'apprentissage profond (AP), que son utilisation a explosé. L'AP

correspond à un ensemble de méthodes d'apprentissage automatique qui visent à modéliser des représentations de données de haut niveau d'abstraction. Ces représentations sont obtenues au moyen d'une succession de couches non linéaires dont l'enchaînement permet d'aboutir à des modélisations complexes (LeCun *et al.*, 2015). L'utilisation de l'AP a permis des progrès spectaculaires en reconnaissance d'images, permettant d'identifier des objets de toutes sortes même si ces derniers avaient des couleurs, des tailles voire des formes différentes (Russakovsky *et al.*, 2015). Ce sont notamment ces méthodes qui se trouvent derrière les outils d'aide au diagnostic médical auxquels nous faisons référence précédemment. Au début de ce projet, ces méthodes commençaient à être appliquées au domaine de l'agriculture (Kamilaris et Prenafeta-Boldú, 2018), et notamment à l'identification de problèmes phytosanitaires (Mohanty *et al.*, 2016; Sladojevic *et al.*, 2016). Ces travaux, qui utilisent des réseaux de neurones convolutifs (RNCs), soulignent le potentiel de l'AP pour la discrimination de maladies du végétal. Les RNCs sont des réseaux de neurones principalement composés d'une succession de couches de convolution et de couches non linéaires qui permettent d'extraire des représentations hiérarchiques des données. Afin de résoudre une tâche, définie au moyen d'une fonction de perte, les valeurs des paramètres internes de ces réseaux sont mises à jour de façon itérative en se basant sur de grandes quantités de données (LeCun *et al.*, 2015). Comme pour les autres techniques d'AP, c'est donc le réseau qui détermine automatiquement la valeur de ses paramètres internes. Du fait des performances obtenues par les RNCs sur des problèmes complexes de traitements d'images, ces réseaux semblent de bons candidats pour relever les défis liés à la prospection automatique. L'association de données proximales acquises au sein de la parcelle et de techniques d'AP pourrait-elle constituer une nouvelle approche pour développer un outil de prospection automatique de la vigne ?

1.3 Objectifs de la recherche

L'objectif général de ce projet est de réaliser une preuve de concept pour un outil de reconnaissance de problèmes phytosanitaires de la vigne basé sur de l'AP. Nous nous concentrons uniquement sur l'aspect analyse automatique d'images d'un tel outil.

Trois objectifs spécifiques visent à :

1. Développer un modèle robuste à la complexité du vignoble, c'est-à-dire aux conditions d'entretien de la vigne et de l'inter-rang, aux conditions météorologiques et d'illumination ainsi qu'à la complexité de l'architecture foliaire et des stades phénologiques.
2. Développer un modèle robuste à la complexité des problèmes phytosanitaires étudiés, soit à la différence d'intensité des symptômes et leur évolution, à la détection de symptômes sur plusieurs organes, à la différence d'expression des symptômes selon

le cépage ou l'année ainsi qu'aux confusions avec des problèmes phytosanitaires aux symptômes proches.

3. Faciliter le développement d'un outil d'analyse d'images opérationnel en explorant des points pouvant s'avérer bloquant pour le déploiement d'un outil de prospection automatique, notamment en étudiant la reproductibilité du modèle d'identification sur plusieurs cépages, sur des images issues de plusieurs capteurs Rouge-Vert-Bleu (RVB) ou plateformes d'acquisition aux résolutions spatiales différentes.

Ces objectifs sont étudiés de manière transversale au sein des trois cas d'étude qui constituent cette thèse.

1.4 Structure de la thèse

Suite à ce premier chapitre introductif, cinq chapitres, dont trois sous forme d'articles, visent à répondre aux objectifs de recherche précédemment formulés.

Dans le Chapitre 2, nous présentons une revue de littérature sur l'utilisation des RNCs pour l'identification automatique des maladies des cultures. Il nous paraissait en effet important de faire un état des lieux sur l'utilisation relativement récente de cette technique d'analyse d'image appliquée à notre domaine d'intérêt et surtout, d'en relever les forces et les faiblesses pour nous permettre de mieux orienter le développement de notre propre méthode. Cette revue de littérature a été publiée dans le journal *Frontiers in Plant Science*, en juillet 2019 : Boulent, J., Foucher, S., Théau, J., St-Charles, P.-L. (2019). *Frontiers in Plant Science*, Technical Advances in Plant Science, 10, 941, doi :10.3389/fpls.2019.00941.

Dans les trois chapitres suivants, trois cas d'étude sur la reconnaissance de problèmes phytosanitaires de la vigne à partir d'images issues de milieux non contrôlés sont successivement présentés. Ce sont à chaque fois des RNCs, parfois transformés en réseaux pleinement convolutifs (RPCs, *Fully Convolutional Network*) lors de l'inférence, qui sont utilisés. L'apprentissage est effectuée de manière supervisée, en apprentissage par transfert avec ré-entraînement complet. Afin de permettre une vision globale des travaux constituant cette thèse, un descriptif de l'ensemble des cas d'étude est proposé dans la Table 1.1. Celui-ci regroupe à la fois le problème phytosanitaire principalement ciblé, les sous-objectifs associés, l'approche sélectionnée et les contributions pour chaque cas d'étude.

Le Chapitre 3 présente un premier cas d'étude sur l'identification automatique du mildiou de la vigne par classification d'image. Le mildiou est une maladie fongique d'intérêt car elle possède plusieurs expressions de symptômes apparaissant aussi bien sur les feuilles que les grappes et les rameaux. Ses expressions évoluent au fil de la saison et selon l'intensité de la contamination. Parvenir à identifier du mildiou suppose donc de regrouper sous une même

classe toute cette variété. Avec des images prises sur deux saisons, sur six cépages et dans trois régions viticoles, nous visons également à évaluer la robustesse du modèle à la diversité des conditions d'entretien des vignobles ainsi qu'à la réalité du champ. Comme ce cas d'étude constitue notre première application des RNCs, nous avons choisi de débiter avec une tâche relativement simple : la classification de petites images par RNCs. Insuffisante pour développer un outil de prospection automatique, elle constitue néanmoins un bon point de départ à l'exploration de la robustesse des techniques d'AP à la complexité agricole et phytosanitaire. Ce cas d'étude a été présenté lors de la *12th European Conference on Precision Agriculture*, qui s'est tenue à Montpellier en juillet 2019, puis a été publié dans les actes de congrès : Boulent, J., Beaulieu, M., St-Charles, P.-L., Théau, J., Foucher, S., (2019) Precision agriculture '19. Proceedings of the 12th European Conference on Precision Agriculture, Montpellier, France, Wageningen Academic Publishers, pp. 141-148, doi : 10.3920/978-90-8686-888-9_16.

Le Chapitre 4 présente un second cas d'étude sur l'identification automatique de symptômes de flavescente dorée (FD). L'intérêt d'étudier la FD réside notamment dans le fait que les symptômes exprimés par une vigne contaminée sont différents d'un cépage à l'autre, voire d'une année l'autre. On se retrouve donc à devoir traiter toute une gamme d'expressions telles que des feuilles plus ou moins enroulées ou encore plus ou moins colorées. Ce spectre de variations entraîne un plus grand risque de confusions avec des maladies aux symptômes visuellement proches. Il complexifie également la tâche d'acquisition des données : dresser un portrait exhaustif des symptômes pouvant être exprimés par la FD est très difficile, et assurément impossible à réaliser sur une unique saison, ce qui pourrait s'avérer être un frein pour le déploiement d'outils de prospection automatique. Nous nous sommes donc intéressés à la possibilité de développer un modèle multi-cépage. Est-il nécessaire de renseigner l'ensemble du spectre des variations de symptômes pour que le modèle fonctionne ou quelques expressions clefs peuvent-elles suffire ? Pour évaluer cet aspect, un RNC a été entraîné sur un cépage symptomatique et testé sur plusieurs autres. Afin de mieux analyser la capacité d'application multi-cépage du modèle, nous avons recours à des techniques de visualisation, ce qui nous permet également de conférer une plus grande transparence à nos résultats d'expérimentation. Enfin, pour nous rapprocher d'un résultat adapté au développement d'un outil de prospection automatique, nous avons utilisé une technique nous permettant d'obtenir une segmentation à partir d'un RNC : la transformation en RPC. Ce cas d'étude a été publié dans le journal *Frontiers in Artificial Intelligence – AI in Food, Agriculture and Water*, en novembre 2020 : Boulent, J., Foucher, S., St-Charles, P.-L., Théau, J. (2020) Automatic detection of Flavescente dorée symptoms across white grapevine varieties using deep learning, doi : 10.3389/frai.2020.564878.

Le Chapitre 5 présente la dernière étude de cette thèse, qui porte sur les performances des




RNCs face à des images acquises avec des paramètres différents de celles sur lesquelles les modèles ont été entraînés. L'intérêt d'étudier cette variation de performance réside dans le fait qu'exiger des paramètres d'acquisition d'images identiques pour toutes les prospections pourrait être un frein au déploiement de telles technologies. En effet, selon la configuration d'une parcelle, son relief ou sa superficie, ainsi que selon les moyens matériels et financiers des agriculteurs, les paramètres d'acquisition idéaux peuvent être amenés à varier. Même si l'objet d'étude reste identique, comme dans le cas de la diversité des expressions de FD selon le cépage, une différence dans les paramètres d'acquisition peut conduire à une divergence importante entre les statistiques des images d'entraînement et celles sur lesquelles on cherche à effectuer l'inférence. Le modèle n'est alors pas capable d'analyser ces images. Dans cette étude, nous nous sommes intéressés à deux paramètres : la résolution spatiale et la plateforme d'acquisition. Nous avons étudié l'impact d'une dégradation de la dimension des images en inférence sur les performances des RNCs. Par cela, nous avons cherché à évaluer si ces modèles (entraînés sur des images acquises par perche) étaient suffisamment confrontés à des objets de dimension restreinte en entraînement afin d'être capables d'analyser des images plus faiblement résolues, voire issues d'acquisition par drone.

Enfin, le Chapitre 6 présente une discussion générale et une conclusion, amenant des pistes de réflexion et des perspectives sur le développement d'outils de prospection automatique pour la viticulture et plus largement en agriculture.

Deux annexes viennent clôturer cette thèse. L'annexe A propose une description générale de l'ensemble des données acquises dans le cadre de ce projet. L'annexe B décrit, quant à elle, quelques éléments techniques, tels que l'équipement informatique utilisé, les bibliothèques employées ainsi que des informations sur la durée d'entraînement et d'inférence des modèles développés.

TABLE 1.1 – Descriptif des trois cas d'étude présentés dans cette thèse.

Pour rappel : OBJ1 – Développer un modèle robuste à la complexité du vignoble ; OBJ2 – Développer un modèle robuste à la complexité des problèmes phytosanitaires ; OBJ3 – Faciliter le développement d'un outil d'analyse d'images opérationnel en explorant des points pouvant s'avérer bloquant pour le déploiement d'un outil de prospection automatique. *Abréviations : FD = Flavescence dorée, RNC = Réseau de neurones convolutifs, RPCs = Réseaux pleinement convolutifs, RVB = Rouge-Vert-Bleu.

	Problème phytosanitaire (PP)	Sous-objectifs visés	Approche	Contributions
Chapitre 3 : Reconnaissance du mildiou par classification d'image	Le mildiou 	<ul style="list-style-type: none"> ❖ OBJ1 : milieu non contrôlé, diversité des conditions d'illumination, des stades phénologiques. ❖ OBJ2 : diversité des PP, des symptômes d'un même PP, des végétaux, des cépages. ❖ OBJ3 : deux capteurs RVB*. 	<ul style="list-style-type: none"> ❖ Classification multi-classe pour la reconnaissance du mildiou et de 4 autres PP. Analyse de portions d'images de 224 * 224 pixels. ❖ Comparaison de performances pour 5 profondeurs de ResNet. 	<ul style="list-style-type: none"> ❖ Taux de bonnes prédictions >95 % sur des images complexes. ❖ Les RNCs sont assez robustes pour classer des portions d'images acquises en milieu non contrôlé et reflétant la complexité des PP.
Chapitre 4 : Reconnaissance des symptômes de FD* par segmentation d'image	La FD 	<ul style="list-style-type: none"> ❖ OBJ1 : milieu non contrôlé, diversité des conditions d'illumination, complexité de l'architecture foliaire. ❖ OBJ2 : diversité des cépages et expressions d'un même PP. ❖ OBJ3 : deux capteurs RVB, évaluation sur des images issues d'un jeu de données externe. 	<ul style="list-style-type: none"> ❖ Segmentation binaire (FD versus non FD) d'images entières. ❖ Entraînement de classifieur et conversion en RPC* lors de l'inférence ❖ Visualisations pour mieux comprendre les critères de prédiction. 	<ul style="list-style-type: none"> ❖ Taux de bonnes prédictions >98% sur des images entières de Chardonnay. Performances plus faibles voire nulles sur d'autres cépages blancs, selon l'intensité de l'enroulement. ❖ Capacité d'analyse conservée sur des images provenant d'une source externe.
Chapitre 5 : Les RNCs* face à des changements de résolution spatiale et à des images acquises par drone	La FD 	<ul style="list-style-type: none"> ❖ OBJ3 : deux plateformes d'acquisition (perche et drone), évaluation des modèles sur des images de différentes résolutions spatiales. 	<ul style="list-style-type: none"> ❖ Segmentation binaire (FD versus non FD) d'images entières. ❖ Entraînement de classifieurs avec 3 jeux de données différents pour augmenter la capacité de traitement d'images à faible résolution spatiale et acquises par drone. ❖ Conversion en RPC pour l'évaluation sur 2 ensembles de test : (1) avec des images acquises par perche, (2) avec des images acquises par drone. 	<ul style="list-style-type: none"> ❖ Résolution spatiale comme paramètre d'intérêt à la fois en entraînement et en inférence. Impact important sur les performances. ❖ Trop grandes différences entre les images perche et drone pour permettre une généralisation du modèle perche sur les images drone.

Chapitre 2

Cadre théorique : L'apprentissage profond pour la détection des maladies des cultures

2.1 Présentation de l'article

Les progrès notoires obtenus en traitement d'image au début des années 2010 ont conduit à une rapide adoption de l'AP dans une multitude de domaines d'application. L'agriculture, et plus particulièrement la reconnaissance automatique de problèmes phytosanitaires, sont également concernées. De premières études utilisant l'AP appliqué à cette thématique ont été publiées à partir de 2016. Les taux de classification qui y sont présentés sont plus élevés que ceux obtenus avec les techniques d'analyse d'images utilisées jusqu'alors. Comme nous l'avons souligné précédemment, analyser des images de végétaux et leur associer un diagnostic est cependant loin d'être une tâche simple. La variété d'expression des symptômes, de stades phénologiques, de cultivars, de conditions d'entretien des parcelles ainsi que la complexité foliaire, soumettent à rude épreuve les algorithmes de traitement d'image. Les résultats obtenus avec l'AP semblent prouver que ces techniques sont capables de gérer toute cette complexité, ce qui serait plus que prometteur pour le développement d'outils d'aide à la prospection.

Cependant, même si ces progrès sont incontestables, quelques zones d'ombre peuvent être relevées, remettant parfois en question ces bons résultats, comme notamment l'utilisation de données d'évaluation peu diversifiées, parfois non indépendantes des données d'entraînement, ou encore une absence de transparence sur les caractéristiques des images utilisées par le modèle pour baser sa prédiction. Pourtant, pour que des modèles d'aide au diagnostic puissent être adoptés par les agriculteurs, il est nécessaire de s'assurer de leur fiabilité.

Cette revue de littérature fait le point sur l'ensemble des travaux publiés entre 2016 et la fin de l'année 2018 et portant sur l'identification automatique de maladies des cultures via l'utilisation de RNCs. L'objectif de cette revue était de déterminer les forces et faiblesses des RNCs pour l'identification des maladies des végétaux, mais aussi de définir quels sont les points d'amélioration nécessaires pour aboutir à un outil d'aide à la prospection. Cet article a été publié dans le journal *Frontiers in Plant Science*, en juillet 2019 : Boulent, J., Foucher, S., Théau, J., St-Charles, P.-L. (2019). *Frontiers in Plant Science*, Technical Advances in Plant Science, 10, 941, doi :10.3389/fpls.2019.00941.

2.2 Résumé

L'AP, et en particulier les RNCs, ont permis d'obtenir des progrès majeurs en traitement d'image. Depuis 2016, plusieurs applications visant à identifier automatiquement les maladies des cultures ont été créées en se basant sur ces techniques. Ces applications pourraient être utilisées pour développer des outils contribuant à l'adoption de pratiques phytosanitaires plus durables, comme des outils d'assistance au diagnostic ou encore de prospection automatique.

Pour évaluer le potentiel de ces techniques d’AP pour de telles applications, nous avons étudié un corpus de 19 études qui se basent sur des RNCs pour identifier automatiquement des maladies des cultures. Notre étude permet d’identifier les principaux enjeux et lacunes des travaux dans ce domaine de recherche. Nous fournissons également des lignes directrices pour améliorer l’utilisation des RNCs dans un contexte opérationnel ainsi que quelques orientations pour de futurs travaux de recherche.

2.3 Convolutional Neural Networks for the Automatic Identification of Plant Diseases : A Review

Abstract

Deep learning techniques, and in particular Convolutional Neural Networks (CNNs), have led to significant progress in image processing. Since 2016, many applications for the automatic identification of crop diseases have been developed. These applications could serve as a basis for the development of expertise assistance or automatic screening tools. Such tools could contribute to more sustainable agricultural practices and greater food production security. To assess the potential of these networks for such applications, we survey 19 studies that relied on CNNs to automatically identify crop diseases. We describe their profiles, their main implementation aspects and their performance. Our survey allows us to identify the major issues and shortcomings of works in this research area. We also provide guidelines to improve the use of CNNs in operational contexts as well as some directions for future research.

2.3.1 Introduction

Plant health and food safety are closely linked. The Food and Agriculture Organization of the United Nations (FAO) estimates that pests and diseases lead to the loss of 20 to 40 % of global food production, constituting a threat to food security (FAO, 2017). Using pesticides is a way of protecting crops from these infestations and thus preserve yields. Their use has been one of the factors behind the increase in food production since the 1950s, enabling it to meet the needs of a growing population (Cooper et Dobson, 2007). However, the use of such substances is not environmentally harmless. Applying these substances negatively impacts biodiversity, including insect, bird and fish populations, as well as soil, air and water quality (Risebrough, 1986; Gill et Garg, 2014; Goulson, 2014; Sanchez-Bayo et Goka, 2014; Knillmann et Liess, 2019). Their use also constitutes a risk to human health, with both acute and chronic effects (Weisenburger, 1993; Bassil *et al.*, 2007; Kim *et al.*, 2016). However, the quantity

of pesticides used is increasing worldwide, with +78 % of tonnes of active ingredients used between 1990 and 2016 (Food and Agriculture Organization of the United Nation, 2018).

Knowledge of a fields' phytosanitary conditions is a decisive factor in limiting the use of pesticides while protecting harvests. Indeed, it enables farmers to carry out proper practices in the right place and at the right time. However, assessing the healthiness of fields is not simple, and it requires a high level of expertise. Indeed, a disease can be expressed differently from one plant species to another, or even from one variety to another. A given symptom may be the result of different problems, and these problems may also combine on the same plant. Even nutritional deficiencies and pests can produce symptoms similar to those of some diseases (Barbedo, 2016). Assessing the healthiness of plots is also time consuming. Checking the condition of each plant several times in a season is not practical on large farms. The difficulty of accessing some crops can also complicate prospection. The automatic identification of diseases by imagery has the potential to solve all these issues by using automatic prospection or expert assistance tools.

The purpose of this article is to synthesize the studies that have used CNNs to automatically identify crop diseases from images and to assess their potential for operational tools. This paper is organized as follows. The research method used to form our analysis corpus is detailed in Section 2. After describing selected studies' profiles, Section 3 presents the main aspects of the implementation of CNN-based methods, as well as their performance. In section 4, we focus on techniques that can help us better understand the trained models in order to avoid the black box effect and to ensure the reliability of the obtained results. Finally, in section 5, we highlight good practices based on both our experience with CNNs and on conclusions from other application areas. Future research directions are also proposed.

2.3.2 Research methodology

The literature search was conducted through SCOPUS for works that matched keywords such as "deep learning", "deep neural network", or "convolutional neural network", along with keywords regarding "diseases", and "plants" or "crops." The references of the selected articles were also checked. Only English-language articles published in established peer-reviewed journals through December 2018 were selected. The search was limited to studies using RGB images and supervised learning. 19 articles met our criteria. Synthesis tables describing these articles are presented in the Annex Section (2.3.7).

2.3.3 Deep Learning Applied to Diseases Identification

Selected Studies' Profile

In the selected corpus, there was a strong interest in market gardening, with tomatoes in 10 of the 19 selected articles. The issue of automatic crop disease identification can be addressed in a general or in a specialized approach. In the general approach (6/19), multi-crop and multi-disease models are trained, while the specialized approach focuses on one crop (13/19). The main similarity of these studies was their focus on analyzing a single organ : the leaves. Only two studies integrated other plant parts (Fuentes *et al.*, 2017, 2018). The selected corpus mentions two application delivery approaches that motivate the development of automatic diseases identification solutions. The first approach is based on the use of mobile expertise tools that provide in-the-field identification capabilities. This approach relies on pictures taken with a regular hand-held camera and centered on leaves. The analysis tools are based on image classification, where a class or category label is assigned to each analyzed image (Figure 2.1A). Ramcharan *et al.* (2017) and Picon *et al.* (2018) implemented mobile applications to use their models in the field. The second application delivery approach is based on automatic phytosanitary monitoring via autonomous vehicles. This is not presented in concrete terms but rather as a development perspective. In addition, to achieve a working model under field conditions, the data used for training must reflect the complexity of the studied environment. Diseases, plants and crops in general are dynamic objects and environments whose appearance can change. Several phenological or symptom development stages, backgrounds, light conditions and even several acquisition scales (one organ, one plant, several plants) must be integrated. It is also necessary to not only identify diseases but also to locate them. Classification therefore does not produce adequate output when used without a localization solution. To obtain their location, two approaches can be used : object detection, which provides identification and location as a bounding box (Figure 2.1B), or segmentation, which provides identification for each pixel in the given image (Figure 2.1C). In the corpus, 16 studies perform classification and 3 use object detection.

Dataset Design

Data Origin and Characteristics The quantity of information and the diversity within the images varies among the studies. Three types of datasets can be defined, depending on their level of complexity (Figure 2.2). The first type consists of images captured under controlled conditions. In this case, images show one leaf picked up in the field and placed on a uniform background, in an environment with controlled illumination (Figure 2.2A). This simplifies the image analysis process by removing any variability related to external conditions

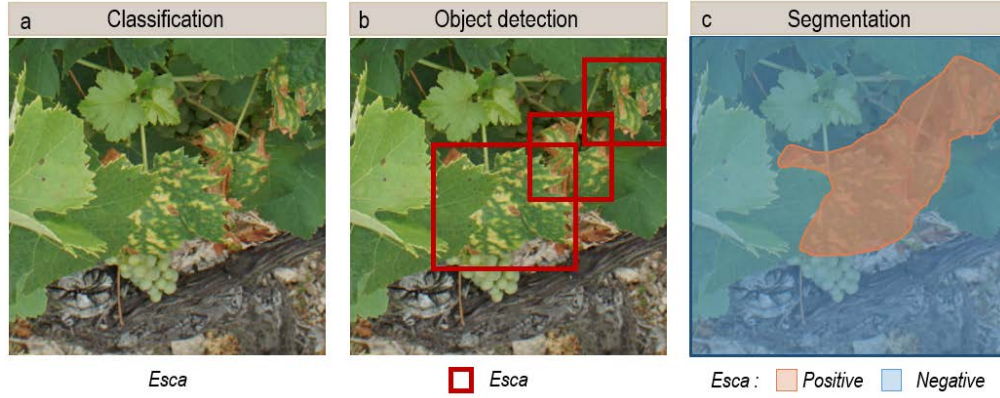


FIGURE 2.1 – Expected output examples of (A) the classification, (B) the object detection, and (C) the segmentation of images containing esca disease symptoms.

or plant morphology in order to focus on symptom expression. A total of 13 of the 19 studies used such images. The second and slightly more complex type of dataset consists of images captured under uncontrolled conditions, but focusing on a particular plant organ, generally a leaf. In this case, images have a complex background but the largest area is occupied by the object of interest (Figure 2.2B). Only 3 of the 19 studies used such images. Finally, the last type of dataset consists of images captured under uncontrolled conditions and without focusing on a particular plant organ. These images therefore reflect what an operator would see in the field, with all the complexity associated with foliage architecture (Figure 2.2C). This kind of dataset is the one best-suited to build an operational automatic phytosanitary monitoring tool. Only 3 of the 19 studies used such images.

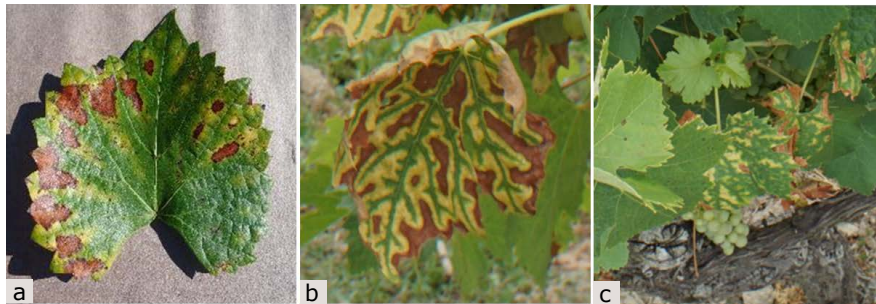


FIGURE 2.2 – Typology of image complexity found in the datasets. Esca grape disease on (A) an image captured under controlled condition (from the PlantVillage dataset), (B) an image captured under uncontrolled condition and with a focus on a particular organ, and (C) an image captured under uncontrolled conditions and without focus on a particular organ.

A majority of the images used in these studies come from public datasets (11 out of 19 studies). The most widely used dataset is PlantVillage, a database initially described in Hughes et Salathé (2015) and now containing 87,848 photographs of leaves of healthy and infected plants

(Ferentinos, 2018). A total of 25 species are represented through 58 classes, with 62.7 % of the images taken under controlled conditions. Barbedo (2018b) used an open database containing 1,383 images of 12 plant species through 56 classes. Its acquisition conditions are mainly controlled. On the other hand, DeChant *et al.* (2017) used a dataset captured under uncontrolled conditions and with no focus on a particular plant organ. This dataset is highly specialized in the identification of Northern Leaf Blight (NLB) infected maize plants. A description of an extended version of this dataset can be found in Wiesner-Hanks *et al.* (2018). It consists of 18,222 annotated handheld, boom and drone images with 105,705 NLB lesions. The availability of such databases is significant, as it provides a high number of annotated images – a key factor of success in deep learning. These images can also be used for benchmarking, enabling a comparison of the accuracy of models created by different research groups. Collecting images in the field and gathering them into databases such as PlantVillage is an ideal solution to improve the community’s research capabilities.

The images in private datasets (the type utilized by 8 studies) were all handheld. A weakness of such datasets is that they sometimes lack important details, such as the capture conditions, acquisition date, varieties studied and the intensity of the analyzed symptoms. In some studies, too few or even no samples are illustrated, making it impossible to determine the application prospects of the trained models. Indeed, a model trained with a low-complexity image dataset (as defined earlier) will not be able to generalize to data from a more complex setting. This was underlined by Mohanty *et al.* (2016), whose accuracy went from 99.35 % on a held-out test set to 31 % on a test set with images taken under different conditions. The work of Ferentinos (2018) divided the PlantVillage dataset based on either laboratory or field conditions. Their model trained on laboratory images reached 33.27 % accuracy when applied on field images, while their model trained on field images reached 65.69 % accuracy when applied on laboratory images. As data is the key element of a successful CNN-based model, its characteristics and source(s) must be well described.

Class Taxonomy Definition Classes are defined by diseases and species (in the case of multi-species models). In Wang *et al.* (2017), classes reflect disease severity levels. Both the intensity and the stages of infection can lead to a high degree of variability in symptoms. This variability can be expressed through separate classes (as did Wang *et al.* (2017)) or integrated in global classes. However, since it is difficult to obtain enough images for all expressions, especially over a single growing season, Barbedo (2018a) recommends the continuous addition of new images to the training dataset. The number of classes vary greatly from one study to another – from 2 (DeChant *et al.*, 2017) to 58 (Ferentinos, 2018). Four studies also dedicated a class to the background. Adding such a class is meaningful for a real world application, where

the background is not uniform. To this end, Brahimi *et al.* (2018) and Sladojevic *et al.* (2016) used the publicly available Stanford Background dataset (Gould *et al.*, 2009). On the other hand, Fuentes *et al.* (2017) extracted patches of healthy plants and background from their images and put them together in a transversal class of negatives. They relied on hard negatives to form this class. Hard negatives are false positives obtained from previous evaluations that are integrated into the training set for the negative class for a new training. This practice aims to reduce the number of false positives by confronting the network to situations that it has previously failed to manage properly. The negative class is quite complex to construct, as it must integrate all the diversity of the real world without having an excessively higher number of images than the positive classes. Using hard negatives can help target the most relevant negative cases. Even if imbalance is a reflection of the field reality where healthy plants are in the majority, some classes should not have a much higher number of examples than others. The class imbalance problem affects both the convergence of the model and its ability to generalize (Buda *et al.*, 2018). Different strategies can be chosen either at a data level or at an algorithm level to minimize the detrimental effect of imbalance (Krawczyk, 2016). Wang *et al.* (2017) chose undersampling. They divided their healthy leaves' class into 12 clusters of 110 images for training and 27 for testing, thereby providing classes of the same size (between 102 and 144 for training and between 23 and 36 for testing). Having a taxonomy that only includes individual diseases is a simplification of reality. Very often, diseases, nutritional disorders and/or pests can be present at the same time, combining their symptoms. Creating a class for each phytosanitary problem combination does not seem to be a suitable solution since the number of possible classes would increase considerably (Barbedo, 2018a). It would be unfeasible to collect enough images for each class. To solve this problem, Barbedo (2018a) proposes to consider lesions individually and process only areas of symptomatic interest identified by the user. Another possibility is to train binary models (target disease versus the rest) and thus give primacy of detection to the disease of interest, even if it does not express "pure" symptoms. In any case, the greatest challenge is to cover enough symptom expressions so that the model can be applied in real world conditions.

Data Annotation Data annotation – the association of a label to all or part of an image – is a laborious but unavoidable step in supervised learning. Repetitive and time-consuming, it must be carried out by an expert in identifying crop diseases, which makes this task difficult to delegate. The annotation method depends on the general approach chosen for image analysis. For classification, it consists of associating a label to each image, either by integrating it in the metadata or by organizing the images, e.g., into folders corresponding to the different classes. For object detection, the coordinates of the target within the image must be entered. This is

done by delineating regions of interest that are often rectangular but may also correspond more precisely to the object in question.

For the identification of diseases, the annotation step raises the question of the analysis scale and of the importance given to the context. Indeed, what is the best piece of information to send to the network? The lesion, the leaf, or the whole plant? Each of these levels are valid and provide complementary features. With a close symptom view, textural elements stand out (Figure 2.3A). A complete view of the leaf reveals symptom patterns (Figure 2.3B). Finally, a view of the whole plant provides a spatial perspective of the symptoms. For example, some problems occur preferentially on young leaves or infect a whole branch (Figure 2.3C, D). Some authors compared those scales and assessed their impact on the final results. Ramcharan *et al.* (2017) formed two datasets : an "original cassava dataset" with entire cassava leaves, and a "leaflet cassava dataset" where the leaflets were manually cropped. The accuracies obtained were slightly higher at the leaflet scale for three of the five studied diseases. Leaf cropping thus had no significant impact, despite the fact that the "dataset leaflet" was seven times wider than the original. Picon *et al.* (2018) addressed the early identification of wheat diseases. Three ways of extracting the tiles sent to the network were tested. The first downsampled the image to match the size required by the network. The second approach was to crop a rectangle containing only the leaf. The third approach, called "superpixel based tile extraction", was based on the segmentation of the image into homogeneous zones. On the test set, the balanced accuracy value was higher at the superpixel scale for one of the two diseases tested. There is therefore no ideal amount of context to give, as it varies from one object of study to another. It is however interesting to note the potential of multi-scale approaches that benefit from information captured at different scales. Note also that a random scaling factor can be added during data augmentation to help models generalize to scale variations (see Section 2.3.3).

Dataset Division When using deep neural networks, three separate datasets are required to develop a model. The first set, the training set, is the collection of images to be used by the network to automatically learn its hidden parameters, such as weights and biases. The second set, the validation set, is used to manually adjust hyperparameters, which are essentially the settings that cannot be automatically learned during training. These include among others the learning rate, the batch size and the network architecture. For more information about hyperparameters, see Goodfellow *et al.* (2016). The values of these hyperparameters are often set empirically, as they are linked to the problem, the dataset and the model architecture. Therefore, there are no good predefined values, as they must be tuned based on the performance (in terms of accuracy) obtained on the validation set. This means that information about the validation data indirectly leaks into the model, resulting in an artificial

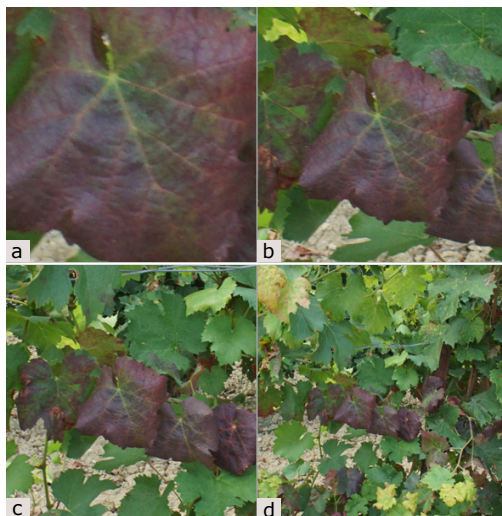


FIGURE 2.3 – Selecting an analysis scale : from a scale close up on the main symptoms (A) to a scale providing more contextual features (D). Example of a vine branch affected by *flavescence dorée*.

ability to perform well on these images (Chollet, 2017). For that reason, the validation images should only be used to tune the hyperparameters; the final evaluation of the model’s performance is done using the test set, discussed in the next paragraph. The model being trained can be evaluated on the validation set at the end of each epoch, allowing the training process to be monitored and to detect overfitting. The training and validation sets come from the same data source that is subdivided. Most of the images go for training (between 70 % and 85 % depending of the size of the dataset). Mohanty *et al.* (2016) and Zhang *et al.* (2018b) tried five different separation ratios and both concluded that using 80 % for training and 20 % for validation was ideal for their data. Another way to divide the images into training and validation sets is cross validation. The dataset is divided into several subsets that are randomly used for training or validation. Cross validation is useful when the dataset is small as it avoids any result bias caused by an arbitrary predetermined and fixed data separation. It is used in 5 of the 19 selected studies. Once the hyperparameters have been defined, further training can be done by gathering the training and validation sets to benefit from a greater number of images.

The third dataset that is needed is the test set. It is used when the training phase has been completed, with the objective of evaluating the model’s final generalization ability. The accuracy on the test set is thus the most important metric to compute, as it provides an overview of the model’s performance beyond the hyperparameter exploration process. The test set must be independent from the training and validation sets, so it cannot be obtained from a simple subdivision. However, 6 of the studies we analyzed (31.5 %) formed their

test set this way. Worse yet, 11 of the studies (58 %) did not even have a test set. Only 2 studies (10.5 %) performed evaluation on an explicitly different test set. Having those three datasets is essential, since the observed data variability in the agricultural setting is quite important. Therefore, there must be a way to ensure that the generated models can operate under different conditions and in different fields. Using data acquired on a different plot for the test set can be a good and simple way to achieve data independence. Even though obtaining data in an agricultural setting can be complex, it should not lead to overlooking this critical aspect.

Data Pre-Processing Before sending images to the network, two pre-processing steps are often necessary. First, the images must typically be resized to match the size of the input layer of the CNN. The sizes are quite standard from one network to another, with for example 227×227 for AlexNet, 224×224 for DenseNet, ResNet and VGG, and 299×299 for Inception. Secondly, the images must be normalized to help the model to converge more quickly as well as to better generalize on unseen data (Chollet, 2017). Other pre-processing operations have been proposed. Oppenheim et Shani (2017) and Mohanty *et al.* (2016) transformed their images to grayscale. Mohanty *et al.* (2016) compared accuracies obtained on grayscale with those from color images. The performance was slightly higher on the color models, with the f1-score improving from 1.34 % to 3.33 % (for details on the f1-score, see Powers, 2011). Even if using color images helps the identification process, as the performance decreases only slightly during the grayscale transformation, this highlights that the network relies mainly on other features to identify diseases. In the same study, the authors also evaluated the impact of background suppression. In fact, background management is one of the challenging elements in the implementation of automatic methods for identifying phytosanitary problems in imagery. With conventional image processing methods, leaf segmentation is a preliminary step to the analysis (Barbedo, 2016). The performance obtained by Mohanty *et al.* (2016) is marginally better with the background, improving the f1-score by slightly less than 1 %. Since background segmentation is not an option on images taken in the field, and since it is the strength of the CNNs to manage complex backgrounds, background suppression is unnecessary.

Training Phase

During the training phase, the model’s internal weights are automatically updated over several iterations. External factors such as the training strategy, architecture, regularization techniques or the value of the hyperparameters influence this training process.

Comparing studies and their results to extract insights on how to define the training phase is

complicated because they do not use the same data and they do not provide all the parameters required to reproduce their experiments. It is also difficult to appreciate the significance of the conclusions made in these studies because their experiments are not performed multiple times to evaluate the impact of random initializations and training sample ordering. Nevertheless, we decided to present some comparisons of training and architectural strategies while keeping in mind that some of their results are potentially biased.

Training Strategies There are two ways to train a CNN : from scratch or with transfer learning. Transfer learning is when a network that is pre-trained on a large set of images (for example ImageNet, and its 1.2 million images in 1000 classes) is used and adapted to another task. This kind of learning is enabled by the fact that the first layers of CNNs learn generic low-level features that are not class specific (Zeiler et Fergus, 2014). In practice, this adaptation is done using the network weights from previous training. Using transfer learning allows us to use CNNs even when the amount of training data is limited, which is often the case in the context of crop diseases identification. This technique helps to achieve greater generalizability, as the network had previously learned to deal with millions of examples. It is also a way to save in terms of computing time and capacity. There are two ways to perform transfer learning : by feature extraction and fine-tuning. Feature extraction consists in keeping the weights of a pre-trained model intact and using the embeddings it produces to train a new classifier on the target dataset. Fine-tuning consists in using the weights of a pre-trained model to initialize the model and then training all or part of these weights on the target dataset (Chollet, 2017). Choosing one technique or the other depends in particular on the proximity between both the source and target datasets (in case they are very close, feature extraction may be sufficient) but also on the size of the target dataset. Training a large number of layers with a small dataset may increase the risk of overfitting. Training from scratch is when the network weights are not inherited from a previous model but are instead randomly initialized. It requires a larger training set, and the overfitting risk is higher since the network has no experience from previous training sessions and so must rely on the input data to define all its weights. However, this approach allows us to define a problem-specific network architecture that can improve the performance. These problem-specific architectures can be developed, for example, to handle more than three color channels, multi-scale dimensions or to integrate multiple models trained differently (with dissimilar hyperparameters or datasets). In our corpus, 15 studies (79 %) used transfer learning and 7 studies (37 %) trained a model from scratch.

Choosing a training strategy depends on both technical (amount of available images, computing capacity) and thematic (availability of a suitable architecture or of pretrained weights

compatible with the data used) considerations. Brahimi *et al.* (2018) compared three training strategies on six CNN architectures (AlexNet, DenseNet-169, Inception v3, ResNet-34, SqueezeNet-1.1 and VGG13). They used the PlantVillage dataset augmented with a background class. Two strategies used transfer learning : feature extraction and complete fine-tuning. In the third strategy, the network was trained from scratch. The accuracies obtained on the validation set and the training time are shown in Figure 2.4. For the six architectures, fine-tuning gave the highest precision (from 99.2 % for SqueezeNet to 99.5 % for VGG13). The times required for fine-tuning and for training from scratch are close (from 1.05 to 5.64h for fine-tuning and from 1.05h to 5.91h when trained from scratch). The feature extraction approach had the lowest training times (from 0.85h to 3.63h).

Overall, training from scratch and transfer-learning should not be seen as completely mutually exclusive strategies. In case of a new architecture where no prior weights are available, pre-training may still provide some benefits. The idea is to start by training the model from scratch on a large dataset (like ImageNet, but also PlantVillage or another plant database) and then carry out fine-tuning on the specific data gathered for the study (Picon *et al.*, 2018; Cruz *et al.*, 2017).

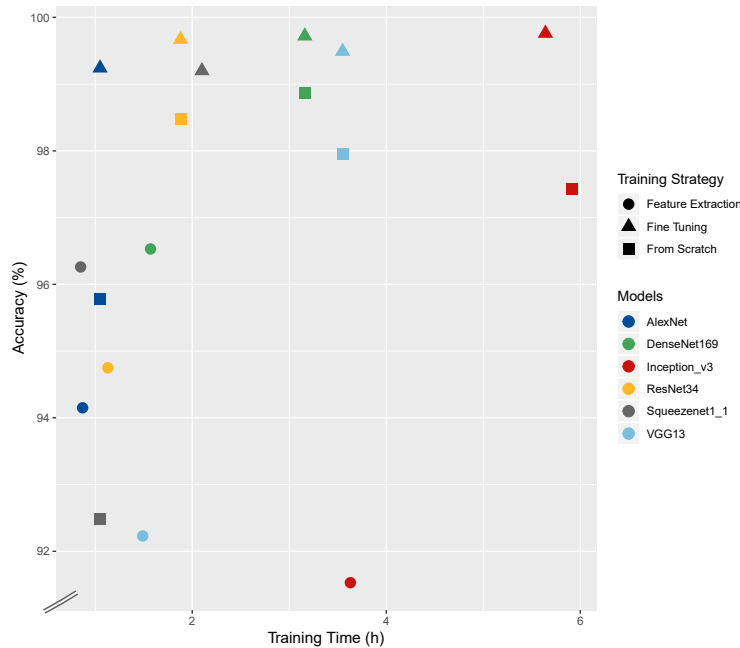


FIGURE 2.4 – Comparison of the training time (h) and accuracy values (%) obtained on a validation set according to different architectures and training strategies. Adapted from Brahimi *et al.* (2018).

Architectures CNNs are based on three main components : convolutional layers, pooling layers and activation functions, commonly Rectified Linear Units (ReLUs). The number of layers used, their arrangement and the introduction of other processing units vary from one architecture to another, determining their specificity. In 17 of the 19 selected studies, state-of-the-art architectures were used. Architectures proposed alongside the first popular CNNs (e.g. LeNet, AlexNet, and CaffeNet) are used in 11 studies. One study used SqueezeNet, known to achieve similar performance as AlexNet on ImageNet with 50 times fewer parameters. The acclaimed ResNet, VGG and Inception architectures are used in 15 studies. Newer architectures such as DenseNet or ResNetXt are used in 3 studies. For more information about architectures, see the work of Khan *et al.* (2020).¹

In 12 of the 19 studies, different architectures were compared, with the highest and lowest accuracy for 4 of these studies reported (Figure 2.5). The accuracies range from 59 % with a ResNet-101 (Fuentes *et al.*, 2017) to 99.75 % with a DenseNet-121 architecture (Too *et al.*, 2018). Network complexity and depth do not necessarily lead to higher accuracy, as shown by the superiority of the VGG-16 results over those of ResNet-101 for Fuentes *et al.* (2017) and the closeness of the results of Inception-V3 and SqueezeNet (Brahimi *et al.*, 2018). The performance reported for a unique architecture may also vary from one study to another, as in the case of VGG-16, which ranked as the best architecture for Fuentes *et al.* (2018) and Wang *et al.* (2017), but as the worst for Too *et al.* (2018). Since architecture implementations are now widely distributed through standard libraries such as PyTorch and Tensorflow/Keras, we advise trying several architectures to find the ones that bring the best results on a studied case. This choice will depend on the nature of the data, its quantity and the time and resources available for training. Choosing an architecture and defining the optimal values of the other hyperparameters can seem like a hazardous trial-and-error process. However, there are methods to guide this process, whether through manual hyperparameters tuning, random or grid research (Goodfellow *et al.* (2016), Bengio (2012), Smith (2018)).

Custom architectures were used in seven studies. In some cases, this involved adapting a reference architecture so that it is more efficient in handling the study data. To limit the risk of overfitting due to their small dataset, Oppenheim et Shani (2017) relied on drop-out, a regularization technique based on the random disconnection of links between model layers. In other cases, architectures were more customized. Cruz *et al.* (2017) implemented what they called an "Abstraction Level Fusion". They injected manually generated features into the fully-connected layers of the network in order to guide its training and possibly accelerate it. Zhang *et al.* (2018b) presented a three-channel convolutional neural network with the goal

1. Dans Boulent et al. 2019, cette référence est associée à une prépublication qui a depuis été publiée. La référence a donc été actualisée.

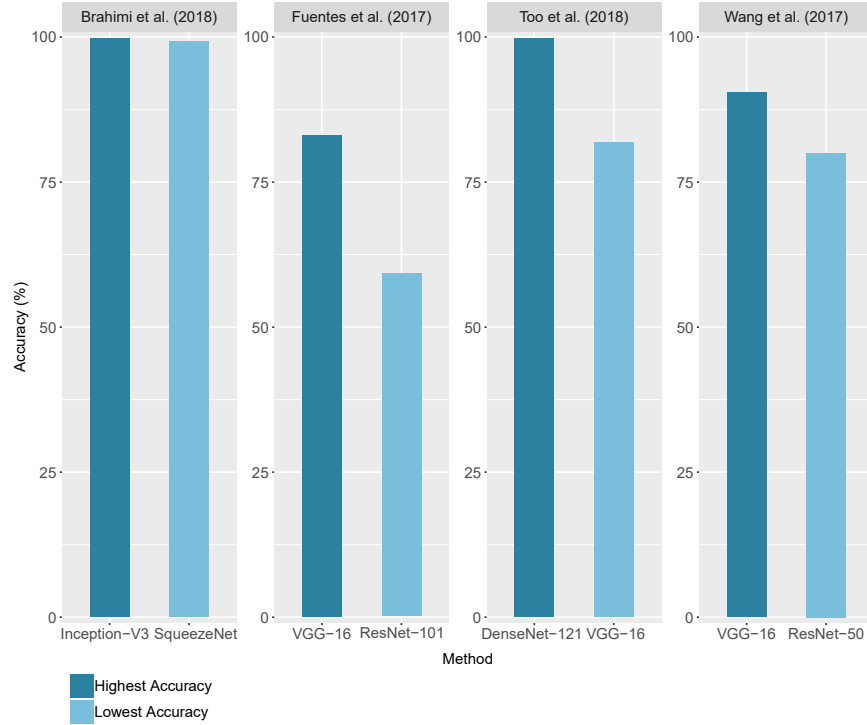


FIGURE 2.5 – Comparison of the accuracies obtained using different architectures. For each study, the lowest and highest performance achieved on the validation set (or test set when available) is reported.

of improving the use of color information. DeChant *et al.* (2017) developed a three-stage process based on the training of several CNNs to compute heat maps which are used to determine if the analyzed image contain diseased lesions. Fuentes *et al.* (2018) implemented a framework based on CNN filter banks to minimize the number of false positives. Customized architectures can be useful to obtain a model more adapted to a specific study case. However, a customized architecture must be compared to a recent state-of-the-art architecture in order to measure its unique contribution.

Regularization Techniques The main challenge in machine learning is to obtain a trained model that is able to analyze new and unseen data. This aspect is far from being guaranteed by high training accuracy. Indeed, the main pitfall in deep learning is overfitting. This occurs when the number of input samples is too small compared to the learning capacity of the network. Overfitting does not allow to learn the general characteristics of the classes and instead captures the noise of the training set (Srivastava *et al.*, 2014). This leads to a model with high accuracy during training but that is unable to generalize (i.e. it does not achieve high test accuracy). In the corpus, trained models are not systematically tested on independent

data (only 2 studies have an explicitly independent test set). It is therefore not possible to determine whether the models were overfitting. However, several of the selected studies presented a low minimum number of samples per class (before augmentation) : equal to or less than 55 for 2 studies, and between 55 and 200 for 8 studies. In those cases, the number of images seems too small to train a model in a robust way - especially considering the diversity in the plant world.

To improve model generalization, the first and obvious step is to gather more data. However, obtaining many images for a given class can be complicated in an agricultural context, especially when it comes to diseases. In machine learning, techniques have been developed to improve the performance of the test set, even if it means reducing the performance of the training set. These techniques are called regularization techniques (Goodfellow *et al.*, 2016). One of these is data augmentation, which consists of the transformation of the geometry or intensity of the original images to make them seem like new images. The operations are often simple : rotation, mirroring, random cropping, zooming or even adding noise, changing contrast and brightness values as well as simulating new backgrounds. The size of the dataset and its diversity are therefore artificially increased. The augmentation operations can be performed in different ways, using one or several (possibly randomly chosen) transformations per image. The transformations can also be applied before the start of the whole training process, or "online" when each image batch is uploaded. An incorrect practice noted in some studies must be highlighted : the augmentation was sometimes carried out before the separation of the images into training and validation sets. It is important to carry out these augmentation operations once the sets have been defined to ensure that an image and its duplicates are in the same set. Additional techniques applied to the model itself, such as drop-out (detailed in Section 2.3.3), or during the training process, such as weight decay and early stopping, were found in the articles. The use of such techniques is recommended.

CNN Performance

Comparison with Other Approaches In image classification, CNNs outperform traditional image processing methods in several applications. This general trend is also observed in the automatic identification of crop diseases. Some of the selected studies compared the performance obtained with CNNs to that of other methods. In all of these studies, the CNN results are better than the others. Figure (2.6) groups the best results obtained for a CNN and for an alternative method in studies that made a comparison. The difference of accuracy ranged from 3 % (Brahimi *et al.*, 2017) to 28.89 % (Liu *et al.*, 2017a).

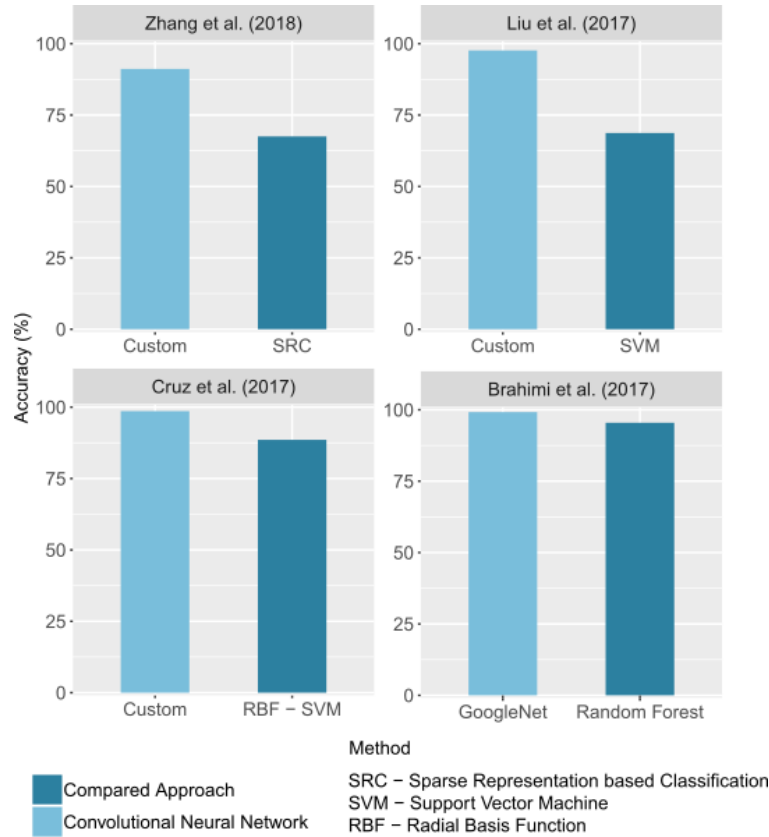


FIGURE 2.6 – Comparison of accuracy values obtained by CNNs and by other image processing methods. Only the best CNN architecture and comparative approach for each study were reported. If given in the study, the precision on a test set was reported.

Generalization Performance To assess the generalizability of a model, it must be evaluated on a dataset of images never seen before, which is the test set mentioned in Section 2.3.3. Only 2 of the 19 studies selected relied on an explicitly independent dataset to conduct this evaluation. Mohanty *et al.* (2016) trained a model to identify 14 crop species and 26 diseases. They used the PlantVillage database. After an accuracy of 99.35 % on a held-out test set, they performed an evaluation on two new datasets of 121 and 119 images, respectively, downloaded from the Internet. On these datasets, they got an accuracy of 31.40 % for the first and 31.69 % for the second. Even if this accuracy is higher than random guessing, it is insufficient for practical use. It can be assumed that the training led to overfitting and/or that the input dataset lacked diversity compared to the test set. Picon *et al.* (2018) aimed to identify three wheat diseases. Using a mobile application, technicians tested the model for two of the three diseases studied : Septoria (77 images) and Rust (54 images), to which they added 27 images of healthy plants. A balanced accuracy of 98 % was obtained for Septoria, and of 96 % for Rust. No drop in performance was observed. This solid performance can be

explained by the fact that the test images were acquired at the end of the season – where the symptoms are most visible – but also thanks to good training practices. Indeed, acquiring more than 1000 samples per class, in several plots in Spain and Germany, over three seasons (2014, 2015 and 2016) and in real environments resulted in enough diversity for the model to work well in operational conditions. These two studies emphasize once again the importance of adopting good training practices and defining an independent test set to properly evaluate the performance of CNNs and their ability to be used operationally.

2.3.4 Understanding the Trained Models

One of the downsides associated with deep learning methods is the difficulty in understanding what the model has learned – the famous black box effect. Still, it is possible to understand the model’s inner mechanism or at least get a glimpse of it through several techniques. By better understanding the trained models, we can not only ensure the relevance of the results generated, but also improve their quality.

Standard Techniques

Easy-to-implement methods that can assess the quality of prediction and gain an insight into how the training progressed have been found in the studied corpus. First, additional metrics to the overall accuracy may be calculated, such as the level of confidence in its prediction and the sensitivity and specificity of the model. By using confusion matrices, inter-class confusion can be assessed. The analysis of incorrect predictions is used to find problematic situations that the model cannot handle. Thus, DeChant *et al.* (2017) observed that illumination variations, background leaves, dead ground vegetation, senescent leaves at the bottom of canopies and insects were badly handled by their model. Fuentes *et al.* (2017) observed confusion between classes with high pattern variations. Ferentinos (2018) pointed out that their model struggles to manage shadow effects, the presence of non-plant objects, and the case where the analyzed leaf occupies a small and non-central part of the image. They may also have found images that were not properly annotated. While these methods lead to a better appreciation of models results and provide clues for improvement, they do not ensure that the training is properly carried out and that the results are due to relevant features. For instance, in our case, these techniques cannot confirm that only plant lesions are used to identify the target diseases instead of other unrelated characteristics from the photos.

Visualization Techniques

To improve the transparency of the learning process, several visualization methods have been developed, allowing us to picture what is happening in the network. Atabay (2017) used the occlusion technique which consists in sliding an occlusion window on the image to study the variation in the probability belonging to the right class. They pointed out that the class was sometimes assigned because of pixels belonging to the background - indicating that the features learned were not just those linked to the symptoms. Brahimi *et al.* (2018) also used the occlusion technique but underlined that it was computationally expensive and time-consuming. They computed saliency maps based on gradient values to estimate the pixels' importance in the node corresponding to the ground truth. They computed this in two ways : with and without guided backpropagation, where only the positive gradients are propagated through the activation functions, which helps to obtain more precise visualizations. Mohanty *et al.* (2016) visualized the top activated feature maps at the output of early convolution layers (Figure 2.7A). Zhang *et al.* (2018a) used t-distributed Stochastic Neighbour Embedding (t-SNE) to visualize the features of their final fully connected layer and to evaluate the distance between their classes (Figure 2.7B). The insight brought by all these visualization methods can help us understand the behavior of trained models while suggesting new improvements. Their implementation minimizes the black box effect, solidifying the reliability that can be attributed to the models, which is decisive for an application in real agricultural conditions. Implementing visualization solutions has also been recommended in the medical image analysis field, where understanding the prediction system is crucial to ensure correct diagnostics (Litjens *et al.*, 2017). For more details on visualization techniques, see Zeiler et Fergus (2014); Qin *et al.* (2018).

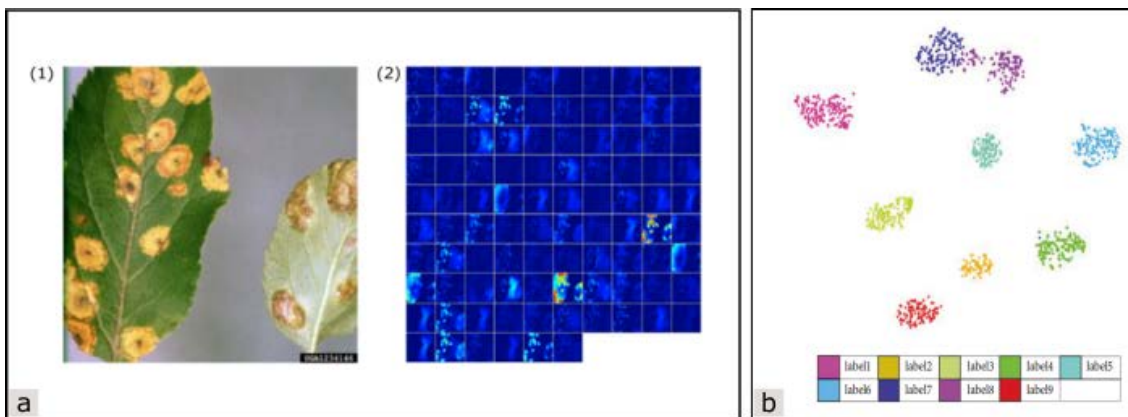


FIGURE 2.7 – Visualization examples found in the corpus (A) Activations in the first convolution layer visualization (Mohanty *et al.*, 2016) (B) T-distributed Stochastic Neighbour Embedding on the final fully connected layer (Zhang *et al.*, 2018a)

2.3.5 Discussion

CNNs provide unparalleled performance in tasks related to the classification and detection of crop diseases. They are able to manage complex issues in difficult imaging conditions. Their robustness may now allow them to emerge from the research environment and become part of operational tools. However, before tools for expertise assistance and automatic screening become a reality, a few steps still need to be tested and integrated. In this section, we first discuss the best practices to adopt all along the development chain so that trained models are able to handle the real-world complexities of agricultural and phytosanitary problems. We then identify the elements to be further addressed to make such tools fully operational, including possible research directions.

Adopting the Best Practices

Targeted Image Acquisition The robustness of a trained model is linked to the quality of its training dataset. Data diversity is one of the key elements to ensure model generalization. Indeed, as also highlighted in Barbedo (2018a), the training dataset has to reflect the reality of the operational environment, which is very challenging. Considering the target application before data acquisition can allow us to capture more appropriate images. For example, for an automatic screening tool, the entire plant must be included in the image, while for an expertise assistance tool, only the leaf or the fruit can be focused on. The operational scope must also be defined in advance. For example, in the case of a tool operating on several varieties or even species, it is necessary to include all the different expressions of symptoms. Carrying out the acquisitions at different hours of the day and under several meteorological conditions ensures the model can operate free from these constraints. Ideally, images should also be acquired on several farms in order to be confronted with a variety of maintenance conditions. Using more than one camera avoids dependency on a particular device. Following this step, the nature of the test set should be considered to ensure that its images will be independent from those used in training. For this purpose, a plot can be devoted to produce the test set. The plot conditions must match the target application scenario, for example by including diseases other than the one studied.

Dataset Preparation The architecture of a model is not the main factor that must be considered to obtain good accuracy. It is rather the quality of the training data as well as its preprocessing and augmentation that can provide the most significant accuracy improvements (Litjens *et al.*, 2017). All steps related to the preparation of the data must therefore be carried out in a rigorous way. The annotation phase should begin by an explicit definition of the class

taxonomy, particularly if the contamination intensity is annotated. This step ensures ensures the annotations reproducibility. Having more than one expert for annotation prevents the risk of dependence on the annotator. Besides, augmentation operations have proven to be effective against overfitting. Easy to set up, they are fully encouraged. It is however imperative to perform these transformations after the separation into training and validation subsets. Otherwise, an image and its transformation could end up in both training and validation sets, causing data leakage. Dealing with the class imbalance is also important for the convergence of the model and its generalizability (see Section 2.3.3).

Training and Evaluation Phases If the time and computing resources allow, conducting several training sessions with the same hyperparameters can lead to improved accuracy, as random initializations can have an impact on the results. When comparing hyperparameters, it would also be advisable to consider fixing the random number generators to prevent them from biasing the comparison. Experimenting with more than one type of architecture can also have a positive effect. For equal accuracy, choosing the least complex architecture is more advantageous from an operational point of view. If applicable, transfer learning is recommended to improve computation time and generalizability. Once all the hyperparameters have been fixed, the model should be retrained by combining the images previously used for training and validation into a global training set. Indeed, as soon as all the hyperparameters have been defined, there is no longer any reason to keep the validation set. It is then worth using this global training set to try to improve the accuracy one last time (i.e., without follow-up adjustments to any hyperparameters). The retrained model can then be evaluated on the test set. The visualization step is also important, as it helps to better understand what is happening in the model and to ensure the robustness of the results. This methodology can also provide opportunities to improve performance.

Sharing Reproducible Results To ensure the results obtained and shared are useful to the scientific community, it is important to have a reproducible research perspective, which is sometimes still lacking in artificial intelligence (Hutson, 2018). When working with CNNs, this requires the sharing of not only all of the hyperparameters, but also data, code and even trained models. This sharing is not always possible because of commercial considerations, but it does allow reproducible works to gain a greater scientific appreciation (Peng, 2011) and to contribute to the progress of the research field.

Outlook

Most applications in crop diseases identification will emerge in uncontrolled environments. Efforts will therefore have to be focused on forming datasets similar to what is found in the field and preferably taken with the acquisition tool of the target scenario application – whether it is a smartphone, a drone, a robot or a tractor. Once new data is acquired, it commonly involves annotation. This step is quite tedious but it is possible to simplify it with active learning. Active learning is an iterative procedure designed to find and annotate the most informative samples. The concept behind this methodology is that annotating good examples can lead to similar or even better accuracy than annotating all examples, and for a lower operative cost as well as requiring less time. A model must first be trained from a small subset of annotated examples. This model is then queried to find the most informative samples (Settles, 2009). Already used as a tool to build datasets in machine learning, this technique is now used with deep neural networks, including for the annotation of medical images (Gorriz *et al.*, 2017; Otálora *et al.*, 2017) or in remote-sensing (Liu *et al.*, 2017b). Another way to reduce the annotation effort would be to share the annotated data. By combining data annotated by different experts, this would also prevent bias related to individual annotators, and it would generally help improve the quality of annotations. Robust and complete datasets could be formed this way. Nevertheless, annotation in such specialized areas is sometimes complex and it can even be difficult for an expert to identify the symptoms of a plant disease. Rigorous annotation is therefore challenging and associated with uncertainty, leading to difficulties in the use of images grouped by social networks (Barbedo, 2018a). Having more benchmark datasets would be very valuable for our field. This recommendation is shared in other areas where annotation requires significant expertise, such as image processing in biology and medicine (Ching *et al.*, 2018). An alternative approach is to avoid the annotation process entirely by using unsupervised learning algorithms for anomaly detection. Already investigated in the medical field (Chalapathy et Chawla, 2019), this could be explored for plant analysis.

In a more thematic way, one other aspect to explore is the early detection of disease. As the first symptoms are more difficult to detect, the use of cameras capturing infrared reflectance would be very interesting. Going beyond the identification model to a decision model is also a perspective. A tool that can identify the problem and provide recommendations for solving it could be a real asset for more sustainable agriculture. The output of the identification model could be just one of the inputs of the decision model, together with auxiliary features such as weather forecasts, geographical characteristics, plot contamination history or disease diffusion pattern.

Next, the classification goal found in most of the corpus’ articles is suitable for expertise

tools but not for automatic screening. Indeed, for this second application, a localization of symptoms is also necessary. In this case, image-space detection or segmentation approaches can be used. Deployment models and user interfaces will also have to be designed. If onboard tools are developed to provide real-time diagnostics, light CNN architectures will be required. The question of minimum spatial resolution will also need to be explored. Close collaboration with farmers could lead to creating solutions fully meeting their needs and financial capacities. User feedback will be decisive, as it will enrich the models with new samples, leading to more robust identification.

2.3.6 Conclusion

In this paper, we identified some of the major issues and shortcomings of works that used CNNs to automatically identify crop diseases. We also provided guidelines and procedures to follow in order to maximize the potential of CNNs deployed in real-world applications. Many already-published solutions based on CNNs are not currently operational for field use mostly due to a lack of conformity to several important concepts of machine learning. This lack of conformity may lead to poor generalization capabilities for unfamiliar data samples and/or imaging conditions, which lowers the practical use of the trained models. Nevertheless, the studied works show the potential of deep learning techniques for crop diseases identification. Their findings are definitely promising for the development of new agricultural tools that could contribute to a more sustainable and secure food production.

2.3.7 Annex

TABLE 2.1 – Summary table – Part 1

Reference	Culture	Training Dataset	Complexity ¹	Number of classes	Number of images	of	Min - Max number samples / class
Atabay (2017)	Tomato	PlantVillage subset	*	10	19742		373 - 5357
Barbedo (2018b)	12 crop plants	Barbedo et al., 2016 (15 % controlled, 85 % in field)	*	56	1383		5 - 77
Brahimi et al. (2017)	Tomato	PlantVillage subset	*	9	14828		325 - 4032
Brahimi et al. (2018)	14 crop species	PlantVillage	*	39	54323		152 - 5507
Cruz et al. (2017)	Olive	Own dataset (controlled)	*	3	299		99 - 100
DeChant et al. (2017)	Maize	Own dataset (field)	***	2	1796		768 - 1028
Ferentinos (2018)	25 crop species	PlantVillage dataset	*	58	87848		43 - 6235
Fuentes et al. (2017)	Tomato	Own dataset (field)	***	10	5000		338 - 18899
Fuentes et al. (2018)	Tomato	Own dataset (field)	***	12	8927		338 - 18899
Liu et al. (2017a)	Apple	Own dataset (controlled and field)	*	4	1053		2366 - 4147
Mohanty et al. (2016)	14 crop species	PlantVillage	*	38	54306		152 - 5507
Oppenheim et Shani (2017)	Potato	Own dataset (controlled)	*	5	400		265 - 738
Picon et al. (2018)	Wheat	Johannes et al., 2017 extended (field)	**	4	8178		1116 - 3338
Ramcharan et al. (2017)	Cassava	Own dataset (field)	**	6	2756		309 - 415
Sladojevic et al. (2016)	Apple, Pear, Cherry, Peach, Grapevine	Own dataset (internet)	NA	15	4483		108 - 1235
Too et al. (2018)	14 crop plants	PlantVillage	*	38	54306		152 - 5507
Wang et al. (2017)	Apple	PlantVillage subset	*	4	2086		145 - 1644
Zhang et al. (2018b)	A/ Tomato B/ Cucum-ber	A/ PlantVillage subset B/ Own dataset (in field)	A/ * B/ **	A/ 8 B/ 5	A/ 15817 B/ 500		A/ 366 - 5350 B/ 100
Zhang et al. (2018a)	Tomato	PlantVillage subset	*	9	5550		405 - 814

¹ NA Not Applicable * Complexity level allowing identification under controlled conditions ** Complexity level allowing identification under uncontrolled conditions *** Complexity level allowing the development of automatic screening tools

TABLE 2.2 – Summary table – Part 2

Reference	Classification or detection ¹	Deep CNN Architecture	Training strategy ²	Best accuracy (%) ³	Evaluation quality ⁴
Atabay (2017)	C	VGG16, 19, custom architecture	FS - TL	97.53	**
Barbedo (2018b)	C	GoogleNet	TL	87	*
Brahimi <i>et al.</i> (2017)	C	AlexNet, GoogleNet	FS - TL	99.18	*
Brahimi <i>et al.</i> (2018)	C	AlexNet, DenseNet169, Inception v3, ResNet34, SqueezeNet1.1.1, VGG13	FS -TL	99.76	*
Cruz <i>et al.</i> (2017)	C	LeNet	TL	98.60	*
DeChant <i>et al.</i> (2017)	D	Custom three stages architecture	FS	96.70	**
Ferentinos (2018)	C	AlexNet, AlexNetOWTBn, GoogleNet, Overfeat, VGG	Unspecified	99.53	*
Fuentes <i>et al.</i> (2017)	D	AlexNet, ZFNet, GoogleNet, VGG16, ResNet50, 101, ResNetXt-101	TL	85.98	**
Fuentes <i>et al.</i> (2018)	D	Custom architecture with Refinement Filter Bank	TL	96.25	**
Liu <i>et al.</i> (2017a)	C	AlexNet, GoogleNet, ResNet 20, VGG 16 and custom architecture	FS -TL	97.62	*
Mohanty <i>et al.</i> (2016)	C	AlexNet, GoogleNet	FS - TL	31	***
Oppenheim et Shani (2017)	C	VGG	Unspecified	96	*
Picon <i>et al.</i> (2018)	C	Custom ResNet50, Resnet50	TL	97	***
Ramcharan <i>et al.</i> (2017)	C	Inception V3	TL	93	**
Sladojevic <i>et al.</i> (2016)	C	CaffeNet	TL	96.3	*
Too <i>et al.</i> (2018)	C	Inception V4, VGG 16, ResNet 50, 101 and 152, DenseNet 121	TL	99.75	**
Wang <i>et al.</i> (2017)	C	VGG16, 19, Inception-V3, ResNet50	TL	90.40	*
Zhang <i>et al.</i> (2018b)	C	Custom Three Channels CNN, DNN, LeNet-5, GoogleNet	FS	A/ 87.15 B/ 91.16	A/ * B/ *
Zhang <i>et al.</i> (2018a)	C	AlexNet, GoogleNet, ResNet	TL	97.28	*

¹ Classification (C) - Detection (D)

² From Scratch (FS) - Transfer Learning (TL)

³ If available, the accuracy of the explicitly different test set is privileged

⁴ * Absence of three explicit subsets ** Three explicit subsets *** Test set explicitly different from the training set

2.3.8 Acknowledgements

This research project was funded by the Centre de Géomatique du Québec (CGQ), Mitacs, the Ministère de l'Economie et de l'Innovation (MEI) of the province of Québec and Microdrones' company.

Chapitre 3

Reconnaissance automatique de maladies de la vigne par classification d'images : Cas d'étude du mildiou

3.1 Présentation de l'article

Les diagnostics automatiques des maladies des cultures obtenus par analyse d'image ont gagné en fiabilité grâce aux RNCs. Même si la majorité des études faisant état de ces progrès utilisent des jeux de données d'une faible complexité, d'autres se sont confrontées à la réalité des champs et ont obtenus des résultats plus que prometteurs. Dans ce premier cas d'étude, portant sur l'identification automatique du mildiou de la vigne, nous cherchons également cette immersion dans la réalité du champ. Pour ce faire, nous avons fait l'acquisition d'images durant deux saisons, dans trois régions viticoles, dans des parcelles plantées en six cépages différents. Nous avons pu intégrer la diversité des stades phénologiques, des conditions d'illumination et d'entretien des vignes. Nous avons également capturé la variété des symptômes et des intensités de contamination au mildiou. Cette maladie est d'ailleurs d'un grand intérêt du point de vue du traitement d'image. Bien que similaires d'un cépage à l'autre, ses symptômes peuvent s'exprimer sur plusieurs organes et vont évoluer au fil de la saison, selon l'intensité de la contamination. Parvenir à détecter le mildiou de la vigne implique de réunir sous une même classe toute cette multiplicité. Nous avons également photographié les autres problèmes phytosanitaires rencontrés. Ainsi, nous avons créé un premier jeu de données riche de diversité. D'un point de vue méthodologique, nous voulions également éviter de reproduire les écueils observés lors de la revue de littérature, notamment en ce qui relève des jeux d'évaluation. En ce sens, lors des terrains de 2018, nous avons isolé des parcelles pour qu'elles servent uniquement à l'acquisition d'images d'évaluation. Par ailleurs, il nous était impossible de réaliser la même stratégie pour le terrain de 2017, les données acquises étant initialement destinées à une autre étude. Afin de tout de même réutiliser les images de ce terrain, nous avons séparé les photos par date, nous assurant ainsi que le même pied à un instant donné n'était pas présent à la fois en entraînement et en évaluation. Cette démarche, bien que n'étant pas idéale, nous a paru être la meilleure solution à notre disposition afin de pouvoir exploiter les images de 2017.

Ce cas d'étude constituant notre première phase exploratoire, nous avons décidé de commencer par suivre une méthodologie classique afin de pouvoir comparer nos résultats à l'état de l'art. A savoir : réaliser une classification mono-label de petites images à partir d'un RNC pré-entraîné avec affinement complet de l'apprentissage. L'architecture de RNC sélectionnée est ResNet, pour laquelle nous avons comparé cinq profondeurs. La profondeur d'une architecture, c'est-à-dire le nombre de couches à paramètres dans le réseau, va définir la complexité des caractéristiques extraites par le RNC. Un réseau plus profond doit ainsi permettre d'obtenir des caractéristiques de plus haut niveau, plus à même de différencier des éléments complexes au sein des images. Ainsi, on suppose qu'une architecture profonde peut permettre d'obtenir

de meilleurs résultats que dans son équivalent à moins de couches. Cependant, en gagnant de la profondeur, un réseau augmente son nombre de paramètres et se retrouve plus à risque de surapprentissage en cas de jeux de données de taille insuffisante. Entraîner plusieurs profondeurs d’une même architecture nous permet d’évaluer si l’une d’elle est plus adaptée à la quantité de données disponibles et à la complexité du problème.

Ce cas d’étude a été présenté lors de la *12th European Conference on Precision Agriculture*, qui s’est tenue à Montpellier en juillet 2019, puis a été publié dans les actes de congrès : Boulent, J., Beaulieu, M., St-Charles, P.-L., Théau, J., Foucher, S., (2019) Precision agriculture ’19. Proceedings of the 12th European Conference on Precision Agriculture, Montpellier, France, Wageningen Academic Publishers, pp. 141-148, doi : 10.3920/978-90-8686-888-9_16.

3.2 Résumé

Depuis le début des années 2010, des progrès conséquents ont été réalisés en classification d’images grâce à l’AP et aux RNCs. Ces avancées permettent de considérer le développement d’outils de prospection automatique des cultures. L’objectif de ce projet est d’identifier automatiquement les symptômes du mildiou de la vigne – *Plasmopara viticola*. Pour évaluer le potentiel de développement d’un prototype commercial, l’acquisition des images nécessaires à cette étude a été réalisée dans des parcelles de plusieurs régions viticoles et plantées de différents cépages. Pour développer notre modèle, l’architecture de RNC ResNet a été sélectionnée. Plusieurs profondeurs de ResNet ont été testées afin de déterminer si l’une est plus adaptée à cette étude et à la quantité de données à disposition. Les précisions obtenues par les cinq profondeurs sont étonnamment proches. Le niveau de bonnes prédictions le plus élevé a été atteint par un ResNet de 18 couches, avec 95.48 % de précision de classification sur un jeu de test indépendant.

3.3 Deep learning for in-field image-based grapevine downy mildew identification

Abstract

Since the early 2010’s, great progress has been made in image classification through deep learning and convolutional neural networks (CNNs). This allows us to consider the development of automatic high-precision scouting tools for agricultural use. The objective of this project is the automatic identification of grapevine downy mildew – *Plasmopara viticola*. To assess the potential of a commercial prototype, several winegrowing regions and grape varieties

were studied. Images were acquired under in-field conditions. The selected CNN architecture used for developing this model was ResNet. Multiple depths of ResNet have been tested to determine which one achieves the best score for this study case and for the amount of data collected. The accuracies reached by five different depths were surprisingly close. The best accuracy is achieved by a ResNet with 18 layers, with 95.48 % on an independent test set.

3.3.1 Introduction

Automatic identification of crop pests and diseases using visible range images is considered to be challenging, especially when it comes to an application under in-field conditions. Barbedo (2016) highlighted some factors making such automatic identification with standards techniques unfeasible under these conditions : 1) the busy backgrounds and the lack of clear border around symptoms making segmentation difficult, 2) the variation of image capture conditions, 3) the fact that the same disease may produce several symptoms and that, conversely, the same symptoms may be caused by different diseases, and 4) the fact that several problems can affect the same plant simultaneously. To address these challenges, innovative image processing techniques must be used. Convolutional neural networks (CNNs) are a type of network used mostly in computer vision problems. They achieve good results in many classification and detection applications and they are now being investigated for diagnostic aid in crop pests and diseases control. Sladojevic *et al.* (2016) and Mohanty *et al.* (2016) were among the first to publish extensive studies on multiple disease detection on different crop types with CNNs. They achieved, respectively, 96.3 % and 99.4 % accuracy rate. However, these performances were obtained on image sets acquired under conditions that do not match those found on the field. In fact, Mohanty *et al.* (2016) reported a drastic drop in performance (31.69 %) while testing images acquired under different conditions compared with those used in training. To help the model to generalize and operate under field conditions, some studies proposed to use image datasets that reflect the complexity of a real agricultural setting. Fuentes *et al.* (2017) trained a model that recognizes nine different types of tomato diseases and pests with an accuracy of 85.98 %. DeChant *et al.* (2017) implemented a system able to identify northern leaf blight infected maize plants with an overall accuracy of 96.70 %. Both authors demonstrated the usefulness of images acquired directly from the field. In a similar fashion, this study uses a set of field images from several phenological stages, disease development levels and grape varieties. The purpose of the set is to train CNNs for identification of grapevine downy mildew – *Plasmopara viticola*. This disease is found throughout the world’s vineyards and leads farmers to use high doses of fungicides to limit its propagation and to avoid heavy crop losses (Ambiaud, 2012). With an automatic tool for downy mildew identification, farmers could have a better understanding of the phytosanitary status of their vineyard and could

therefore set up optimised and more sustainable pesticide management.

Another aspect of this work is the investigation of neural network architectures with an emphasis on the impact of network depth on the classification accuracy. It is assumed that the depth of networks - i.e. the number of layers with parameters - leads to higher level features. In return, these features improve the aim of discriminating between more complex elements and are less sensitive to large irrelevant variation like background content (LeCun *et al.*, 2015). In consequence, deep architectures should give better accuracy compared to shallow ones. However, the number of parameters to estimate gets larger as the network gets deeper and this pushes the network to overfit if insufficient data are used during the training phase. This is a real problem in the context of crop diseases and pests identification because datasets are usually small and unevenly annotated, meaning that some classes can have few labels for training. Also among other things to consider is the operational level of the system to provide farmers with quick diagnosis. In this case, smaller networks may be proven more efficient. Nonetheless, the selection of a viable architecture for real time application depends on both factors : accuracy and speed.

The objective of this work was to determine the optimal network depth for this study case, i.e. the one that provides the highest accuracy on the test set – demonstrating its ability to generalize. There exists a broad variety of architectures to choose from, but the study focuses on the Residual Network (ResNet, He *et al.* (2016)). A performance comparison between the five depths of ResNet for grapevine pest and disease identification was done together with an evaluation of the ability of a single model to work on several grape varieties, in several wine-growing regions, under various acquisition conditions and on a wide range of symptoms.

3.3.2 Materials and Methods

Data acquisition

Images used to train and evaluate the models were acquired in 2017 and 2018, between June and September. The 2017 acquisitions took place in the vineyard of Agriculture and Agri-Food Canada’s experimental farm in Frelighsburg (45°2’60"N, 72°51’36"W), located in the Montérégie wine-growing region of Canada. The acquisitions covered three grape varieties : Chancellor, Geisenheim and Frontenac. The 2018 acquisitions were carried out in these plots but also in two other French wine-growing regions : Cognac and Languedoc-Roussillon. Data were acquired around the cities of Cognac (45°41’24"N, 0°19’12"W), Carcassonne (43°12’36"N, 2°19’48"E) and Limoux (43°2’60"N, 2°12’36"E) in commercial vineyards. The acquisitions covered three grape varieties : Ugni Blanc, Chardonnay and Chenin.

The photos were taken using standard cameras in the field, either by hand or using a 2 m

pole, which allowed a more global and elevated view of the vines. Leaves, stems, grapes and background were included in the photos (without prior arrangement), under various weather conditions and exposures. The only requirement was that the vines should be free of water drops. These acquisitions produced a database of 1192 images.

Dataset

Two separate datasets (for training and testing) were created based on the captured images. These datasets are based on different plots. This geographical separation provided independence between the training and test sets. The training set contained 899 images, from which 80 % were used for training the models and 20 % for validating their performance. The test set contained 293 images from other plots. However, the geographical separation was impossible for images acquired in 2017 in Montérégie. Therefore, a random separation into 80 % for training and 20 % for testing was performed for that subset. The training set contains photos of Ugni Blanc, Chancellor, Geisenheim and Frontenac varieties. The test set contains photos of these four grape varieties plus Chenin and Chardonnay. The proportion of samples for each location and grape variety is described in Table 3.1.

TABLE 3.1 – Samples origin and grape varieties per dataset and class (in %). The percentage distribution between grape varieties in Frelighsburg is broadly estimated.

	Downy Mildew		Japanese Beetle		Deficiencies		Healthy		Vine Leaf Blister		Dry Plant Organ		Background	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Carcassonne - Chardonnay	0	29	0	0	0	51	0	5	0	0	0	33	0	13
Cognac - Ugni Blanc	67	29	0	0	100	4	48	14	100	76	48	29	38	19
Frelighsburg - 90 % Chancellor - 5 % Geisenheim - 5 % Frontenac	33	39	100	100	0	18	52	77	0	0	52	29	62	66
Limoux - Chenin	0	3	0	0	0	27	0	4	0	24	0	9	0	2

The images were then annotated manually by adding points in salient image regions in order to label them according to their local content. Even if the main goal is the identification of downy mildew, seven labels were defined, which include other phytosanitary problems. This was done to avoid confusion between these problems and also to assess the discrimination capacity of the CNNs. Those labels were defined by their plot’s phytosanitary state : Downy Mildew (DM), Japanese Beetle damages (JB, *Popillia japonica*), Deficiencies (D), Healthy (H), Vine Leaf Blister (VLB) caused by the mite *Colomerus vitis*, Dry Plant Organ (DPO) and Background (B). Several deficiencies, including phosphorus, magnesium and iron, were grouped under the same class due to the lack of samples to label them individually. The DM class includes samples from several disease stages (sporulation, oil spots and drying) and levels of contamination.

Samples of each class were created from 318 by 318 pixels regions of interest for each annotated point. The training set was made up of 6055 samples, the validation set of 1513 samples and the test set of 1883 samples (Table 3.2). Figure 3.1 illustrates the intraclass diversity of the 7 classes with some examples for each of them. Data augmentation was performed on the training set with random rotation, flip and zoom transformations. All the samples were resized to 224*224 pixels and normalised.

TABLE 3.2 – Samples per dataset and class.

Label	Train	Validation	Test
Downy Mildew	2038	510	451
Japanese Beetle	330	82	227
Deficiencies	400	100	248
Vine Leaf Blister	142	35	42
Dry Plant Organ	94	23	45
Healthy	2837	709	716
Background	214	54	154
Total	6055	1513	1883

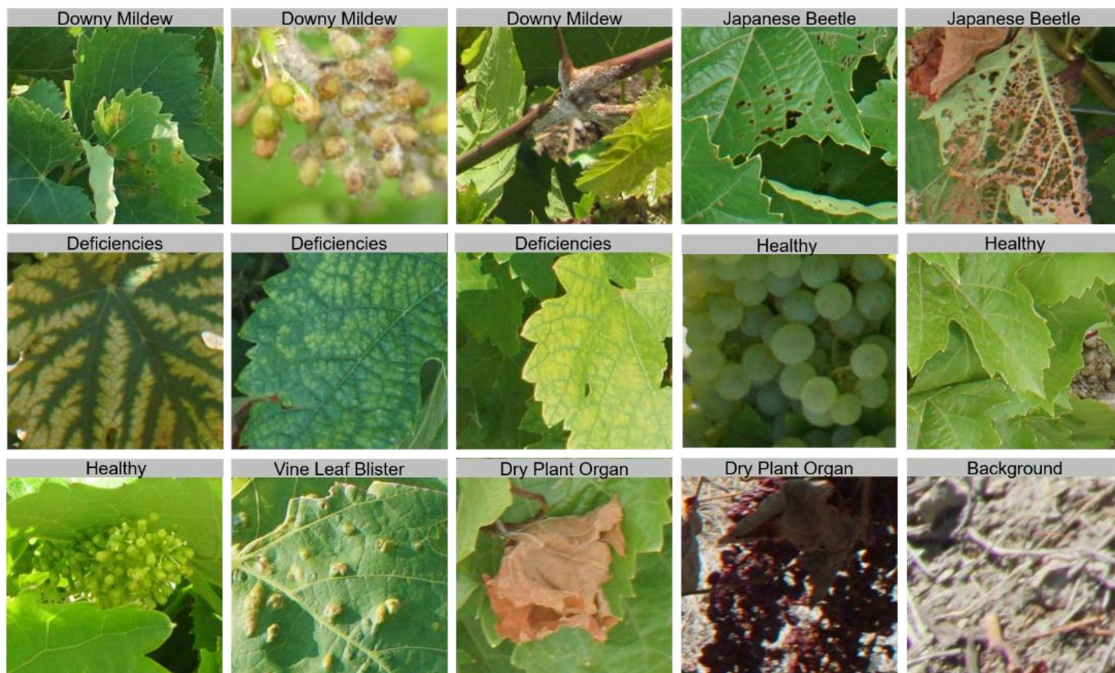


FIGURE 3.1 – Illustration of classes and their diversity

Training and validation phases

This study focused on ResNet networks (He *et al.*, 2016). This is now an established architecture thanks to its performance and fast convergence. The architecture comes in different depth configurations : shorter ones (18 layers), longer ones (34 and 50 layers), and very deep ones (101 and 152 layers). Deep residual learning was introduced to overcome the degradation problem as observed by He *et al.* (2016). The degradation problem corresponds to the saturation and rapid degradation of the training accuracy observed with the increasing network depth. To improve convergence while having very deep architectures, the network integrates shortcut connections. These connections provide an identity map from a given layer to a deeper one and whose value is added to the output of the deeper layer. This allows the intermediate layers of a block to learn a residual function referring to the block input and alleviate the degradation problem. The network’s number of parameters ranges from roughly 11M for ResNet-18, to over 58M for ResNet-152.

To compare their performance on the vine disease classification problem, the five architectures were trained with the same hyperparameters. The learning rate was initialized at 0.0005 and decreased by 90 % every 10 epochs. This small value was selected due to the use of transfer learning. The batch size was 32. The optimizer used was Adam (Kingma et Ba, 2015). The networks were trained for 30 epochs. The training phase was initialized with ImageNet pre-trained weights to allow transfer learning. The goal of transfer learning is to use knowledge learned from tasks for which a large amount of labelled data is available to improve performance in settings where less data is available. This is possible because of the hierarchical nature of the CNNs features. The first layers represent general features, likes corners, edges and colour combinations. The deeper layers represent complex and class-specific structures (Zeiler et Fergus, 2014). The features from the first layers then can be used to describe any object of study. And a partial or total re-training enables the weights to be fine-tuned for the target dataset. This technique reduces processing time and leads to more generalizable models. In this study, all layers were fine-tuned.

Each model was trained five times to ensure the reproducibility of the results. For each training session, the seeds determining the initial state of the random number generators have been fixed. Random values were used among other things here to determine the ordering of the training samples and the augmentation operations to apply to these samples. Five seeds have been fixed with distinct arbitrary values. Each network was trained successively with these five seeds. This guarantees that the performance differences are due to the networks depth differences and not to random initialization collateral effects. For each ResNet, the seed that provides the best overall accuracy on the validation set was kept, and the five best models were then evaluated and compared with the test set.

The processing was implemented with a publicly available deep learning framework¹ based on PyTorch2² developed at the Computer Research Institute of Montréal. Figure 3.2 gives an overview of the processing setup.

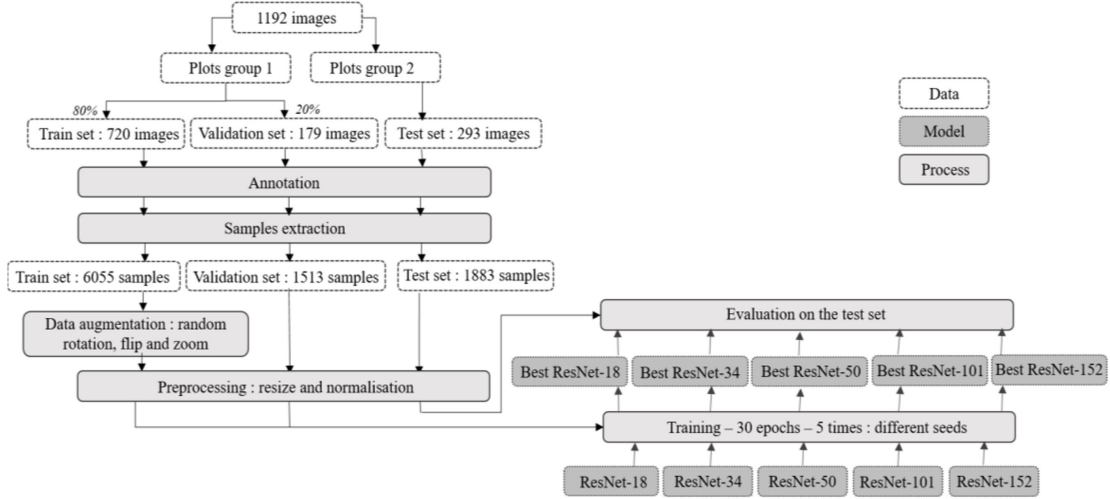


FIGURE 3.2 – Processing setup overview

3.3.3 Results and Discussion

The 25 training sessions resulted in similar and high validation accuracies, fluctuating from 97.28 % to 98.48 %. The same seeds did not produce the best model for the five depths. The best epoch value for each depth according to the five different seeds varied from 0.46 % to 0.96 %, demonstrating the importance of running several training sessions. The best epoch value on the validation set for the five depths varied from 98.11 % for ResNet-34 to 98.48 % for ResNet-18. A larger disparity occurred on the accuracy achieved on the test set, reaching from 93.73 % for ResNet-101 to 95.48 % for ResNet-18 (Table 3.3). The ResNet-18 achieved the best results. This performance remains relative to this study case and to the parameters specified for training. Indeed, the five depths produced almost similar performance. Given the marginal differences between the results, different seeds could lead to a different ranking. However, the good performance of a low depth network indicates that low-level features are sufficient for this classification task. Adding more layers did not result in better features or discrimination. These findings are very promising on an operational level, as a shallow model allows faster executions with limited computing power.

The classification report shows that three classes had very good performance scores (F1-Score

1. Link to the framework used : <https://github.com/plstcharles/thelper>

2. <https://pytorch.org/>

TABLE 3.3 – Accuracy of the best models for each ResNet (%).

	ResNet-18	ResNet-34	ResNet-50	ResNet-101	ResNet-152
Best model epoch	30	28	28	24	29
On validation set (%)	98.48	98.11	98.24	98.14	98.27
On test set (%)	95.48	94.47	95.43	93.73	94.10

< 96 %) : DM, JB and H (Table 3.4). The model performed less well for two of the seven classes : VLB and D, with a respective F1-Score of 88.00 % and 89.08 %. For VLB, this can be explained by the small number of samples for training. Also, 14.29 % of misclassified VLB were classified as H, the class with the biggest number of false positives (Figure 3.3). This amount of false positive in H may be due to the diversity covered by this class. In fact, depending on the grape varieties and phenological stages, the leaves can have colours ranging from yellow-green to dark green, with undulating to smooth textures. A healthy leaf and a leaf in early stage of VLB contamination have close characteristics, both in terms of colour and texture which could explain in part the observed confusion (Figure 3.4.a). For D class, there is a lack of samples for each deficiency in its different stages. In fact, it was found that deficiencies samples with intense and advanced symptoms were present only in the test set. Therefore, the model had never been confronted with samples with similar characteristics. From the 36 misclassified samples of class D, 23 were advanced symptoms (Figure 3.4.b and c). DM class achieved a F1-Score of 96.09 %. Some false negatives can be explained by spots with similar patterns found on D class (Figure 3.4.d). The capacity to explain some of the errors gives the positive sense that they can be corrected, in part by adding more images covering those more complex cases.

TABLE 3.4 – Classification report from ResNet-18 evaluation on the test set.

	Precision (%)	Recall (%)	F1-Score (%)	Support Count
Downy Mildew	96.85	95.34	96.09	451
Japanese Beetle	99.53	94.27	96.83	227
Deficiencies	92.98	85.48	89.08	248
Healthy	95.32	99.58	97.40	716
Vine Leaf Blister	100.00	78.57	88.00	42
Dry Plant Organ	93.48	95.56	94.51	45
Background	90.53	99.35	94.74	154
Weighted average	95.56	95.49	95.43	1883

The model was also able to process unknown grape varieties. Chardonnay and Chenin were only shown in the evaluation phase. 12.8 % of Chardonnay samples were misclassified but

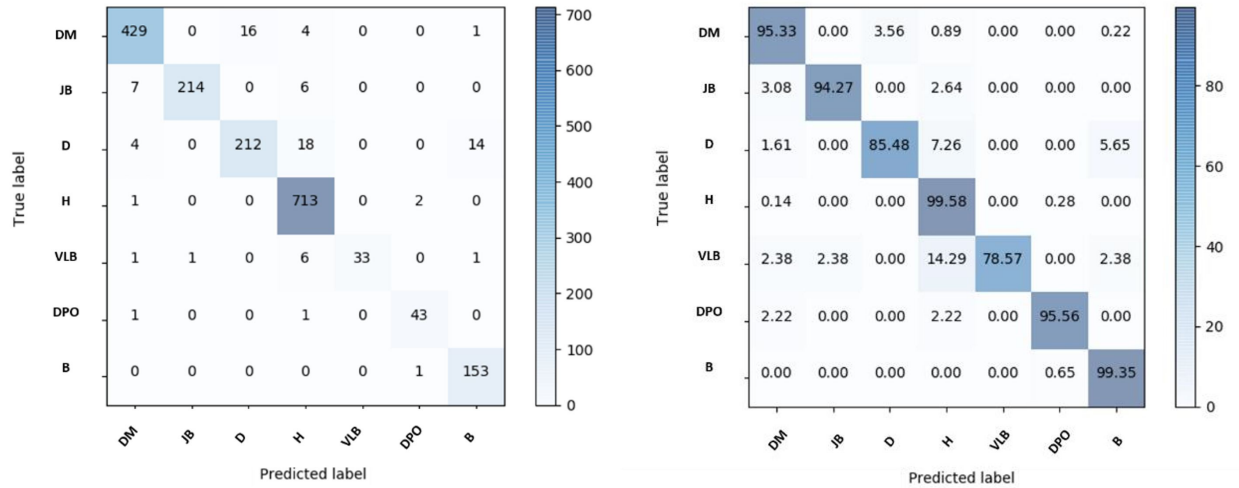


FIGURE 3.3 – Original (left) and normalised (right) confusion matrices (%) from ResNet18 evaluation on the test set

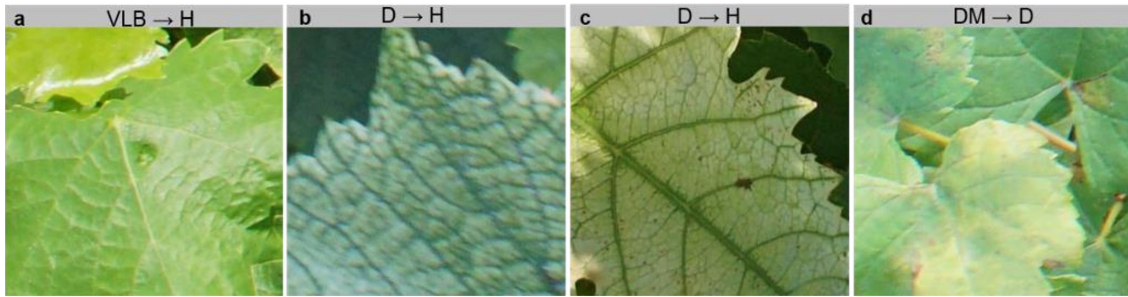


FIGURE 3.4 – Examples of wrong predictions. True Label \Rightarrow Predicted Label. (VLB : Vine Leaf Blister, DM : Downy Mildew, D : Deficiencies, H : Healthy)

59 % of these errors correspond to those presented in Figure 3.4.b and c. As for Chenin samples, only 3.14 % were misclassified.

Finally, only 3.52 % of the misclassified samples were not leaves. Even if leaves represent the vast majority of annotations, this underlines the capacity of the model to process all the vine's organs and different backgrounds.

3.3.4 Conclusion

In this study, through the CNNs, most of the challenges making the automatic identification of pests and diseases unsuitable under in-field conditions (see Introduction) seem to have been handled. Due to CNNs ability to manage busy backgrounds and foliage complexity, prior segmentation was not necessary and images of unarranged plants were processed. Next, the images collected over several days, at different hours and under different lighting conditions and poses were all processed by this model. The classification errors did not seem to be linked

to these factors. The variety of disease symptoms over all plant organs could also be managed – the variety of grapevine downy mildew symptoms being a good example to assess this aspect. Going even further, this study demonstrates that CNNs can identify diseases and pests on several grape varieties, with different vine training systems, at different phenological stages. The model produced here is not yet fully operational. In particular, it should be adapted to process complete images and not just pre-selected samples. It should also handle cases where a single plant may be affected with multiple phytosanitary problems. Given the advances in other related fields, CNNs will undoubtedly be able to meet these challenges. The fact that networks with low depth such as ResNet-18 produce such good results is very encouraging for operational deployment. CNNs are undeniably a technique that will allow the commercial development of tools for the automatic identification of pests and crop diseases. With such tools, it can be expected that scouting will be less painful for farmers and that they will be able to use it to change their phytosanitary practices to more sustainable and environmentally friendly ones.

3.3.5 Acknowledgements

This research project was funded by the Centre de Géomatique du Québec (CGQ), Microdrones and Mitacs. We also thank the Ministère de l’Economie, de la Science et de l’Innovation (MESI) of the province of Québec, Agriculture and Agri-Food Canada, the Chambre d’Agriculture de l’Aude, the Fédération Départementale de défense contre les Organismes Nuisibles (FEDON) de l’Aude as well as all the owners and wine growers who gave us access to their vineyards and/or guided us in the image acquisition process.

Chapitre 4

Reconnaissance automatique de maladies de la vigne par segmentation d'images : Cas d'étude de la flavescence dorée

4.1 Présentation de l'article

Le cas d'étude sur l'identification du mildiou montre que les RNCs sont capables de gérer la complexité inhérente des acquisitions en champ mais aussi de gérer la complexité de l'expression des maladies. Ces résultats confirment le choix de cette technique d'analyse d'image pour le développement d'un outil de prospection automatique, mais sont insuffisants en eux-mêmes pour développer un tel outil. En effet, en plus d'une diminution nécessaire des confusions entre problèmes phytosanitaires mais aussi une diminution des faux positifs au niveau du sol ou de l'arrière-plan, d'autres problèmes techniques et thématiques doivent être résolus. Trois d'entre eux sont étudiés dans ce cas d'étude portant sur l'identification de symptômes d'une autre maladie de la vigne, la FD.

Premièrement, la classification générée dans le chapitre précédent a permis d'obtenir la prédiction d'un label unique pour une image donnée, ce qui n'est pas une prédiction adaptée à notre cas d'application. En effet, être alerté de la présence de la maladie recherchée dans une image n'est pas une information suffisante. Avoir une localisation des zones contaminées constituerait une information plus complète et permettrait également d'augmenter la transparence et donc la fiabilité de la prédiction. En cela, la segmentation semble une approche plus appropriée que la classification. Cependant, la mise en place d'un modèle de segmentation s'avère être plus exigeant en terme d'annotation puisque des masques des classes doivent être envoyés au réseau lors de l'entraînement. Pour ne pas complexifier l'étape d'annotation, nous nous tournons vers les RPCs. Ces réseaux permettent d'obtenir une carte de segmentation à la résolution grossière par le simple remplacement de la couche pleinement connectée d'un modèle de classification en une couche convolutive 1x1. Ce niveau de résolution spatiale grossier correspond bien à notre cas d'étude puisque la FD s'exprime sur toute la feuille et contamine l'ensemble d'un rameau, si ce n'est le pied au complet. Nous avons donc testé cette technique de conversion d'un RNC en RPC, peu utilisée dans le domaine de l'identification automatique des problèmes phytosanitaires.

Le second problème sur lequel nous nous sommes penchés est le niveau de fiabilité trop insuffisant des modèles de détection automatique pour être utilisés par des agriculteurs. L'utilisation de la segmentation au lieu de la classification est déjà une étape vers plus de transparence : il est possible de visualiser en un coup d'oeil si les zones détectées par le modèle correspondent effectivement à des zones symptomatiques. Cependant, nous souhaitons aller plus loin en nous assurant que ces zones sont détectées pour les bonnes raisons. Pour ce faire, nous utilisons une technique de visualisation, le *Gradient-weighted Class Activation Mapping*. Cette technique permet de visualiser pour une image donnée les caractéristiques qui sont utilisées par le modèle pour réaliser sa prédiction. Grâce à cela, il est possible non seulement

de vérifier que le diagnostic se base sur des caractéristiques pertinentes, mais également d'identifier les atouts et limites du modèle, ce qui est crucial dans le cadre d'une application commerciale.

Enfin, le dernier point exploré au travers de ce cas d'étude est l'interrogation sur la capacité d'un modèle à détecter une maladie pour lesquelles toutes les nuances de symptômes ne sont pas renseignées en entraînement. Sur ce thème, la FD est un bon sujet d'étude. Ses symptômes s'expriment en effet différemment d'un cépage à l'autre voire d'une année à l'autre. On retrouve évidemment des caractéristiques communes : rougeoiement des feuilles pour les cépages rouges, jaunissement des feuilles pour les cépages blancs, enroulement de la feuille, non aoûtement du bois et absence de baies. Mais ces symptômes vont s'exprimer dans des intensités variables, ce qui complexifie le diagnostic en augmentant le risque de confusion avec des problèmes phytosanitaires aux symptômes proches, mais aussi en compliquant la tâche de constitution de la banque d'images d'entraînement. Pour renseigner toutes les expressions de FD, des acquisitions pluri-cépages sur plusieurs saisons seraient nécessaires, ce qui pourrait être un frein au développement d'outils de prospection. Est-il nécessaire de renseigner toutes les expressions d'une maladie ou ses symptômes les plus typiques suffisent ? Pour explorer cet aspect, notre idée initiale était d'entraîner des modèles généralistes (plusieurs cépages) et spécialistes (mono-cépages) afin de comparer les résultats obtenus sur différents cépages. Cependant, nous avons été limités par la réalité de notre terrain d'étude. Malgré des prospections sur plusieurs parcelles d'une dizaine de cépages différents, nous n'avons pu observer assez de contamination sur un seul cépage : le Chardonnay. Pour explorer l'aspect multi-cépages, nous avons choisi d'entraîner un modèle sur des symptômes de FD exprimés sur le Chardonnay et d'évaluer sa capacité de détection sur d'autres cépages blancs, notamment sur des photos fournies par un technicien de la Fédération Régionale de Défense contre les Organismes Nuisibles (FREDON) d'Aquitaine. Ce petit jeu de données possède le grand intérêt d'être issu d'une acquisition complètement extérieure à ce projet et nous permet donc de tester la robustesse du modèle à des images prises dans des conditions d'acquisition inconnues ou non maîtrisées (appareil photo, conditions météorologiques et point de vue). Cet article a été publié dans le journal *Frontiers in Artificial Intelligence – AI in Food, Agriculture and Water*, en novembre 2020 : Boulent, J., Foucher, S., St-Charles, P.-L., Théau, J. (2020) Automatic detection of Flavescence dorée symptoms across white grapevine varieties using deep learning, doi : 10.3389/frai.2020.564878.

4.2 Résumé

La FD est une maladie de la vigne causée par des phytoplasmes et transmise par des cicadelles. Elle génère beaucoup d’inquiétudes dans le secteur viticole du fait de son caractère irréversible et de sa diffusion rapide dans les vignobles européens. Dans cette étude, nous visons le développement d’un outil de reconnaissance automatique de la FD, qui pourrait par la suite être déployé dans les vignobles pour réaliser des prospections automatiques. Grâce à un tel outil, un prélèvement et une analyse laboratoire d’échantillons des pieds possiblement contaminés peuvent être conduits rapidement. En cas de diagnostic positif, la vigne peut ainsi être arrachée sans délais afin d’éviter de nouvelles contaminations au sein de la parcelle. Pour développer un tel outil, nous nous sommes servis de techniques d’apprentissage profond. Plus particulièrement, des RNCs ont été entraînés sur des portions d’images, puis ont été convertis en RPCs lors de l’inférence. Cette transformation permet d’obtenir comme prédiction une segmentation grossière, tout en ayant seulement entraîné un classifieur, moins exigeant en termes d’annotations. Le modèle, entraîné sur le cépage blanc Chardonnay, a été testé sur cinq autres cépages blancs aux expressions de symptômes plus ou moins éloignées. Sur les deux plus larges ensembles d’inférence, le taux de vrais positifs s’élève à 98,48 % pour les photos de Chardonnay, mais tombe à 8,3 % pour les photos d’Ugni-Blanc, ce qui souligne le besoin d’un entraînement multi-variétal afin de saisir la diversité des symptômes de FD. Pour mieux comprendre le fonctionnement et la sensibilité de notre modèle, nous utilisons ensuite deux techniques de visualisation : *Guided Grad-Cam* et UMAP. En plus d’apporter une plus grande transparence à nos résultats, elles nous permettent également d’identifier des pistes d’amélioration afin d’aboutir à la création d’un outil de prospection au service des viticulteurs.

4.3 Automatic Detection of Flavescence dorée Symptoms across White Grapevines Varieties using Deep Learning

Abstract

Flavescence dorée (FD) is a grapevine disease caused by phytoplasmas and transmitted by leafhoppers that has been spreading in European vineyards despite significant efforts to control it. In this study, we aim to develop a model for the automatic detection of FD-like symptoms (which encompass other grapevine yellows symptoms). The concept is to detect likely FD-

affected grapevines so that samples can be removed for FD laboratory identification, followed by uprooting if they test positive, all to be conducted quickly and without omission, thus avoiding further contamination in the fields. Developing FD-like symptoms detection models is not simple, as it requires dealing with the complexity of field conditions and FD symptoms' expression. To address these challenges, we use deep learning, which has already been proven effective in similar contexts. More specifically, we train a Convolutional Neural Network (CNN) on image patches, and convert it into a Fully Convolutional Network (FCN) to perform inference. As a result, we obtain a coarse segmentation of the likely FD-affected areas while having only trained a classifier, which is less demanding in terms of annotations. We evaluate the performance of our model trained on a white grape variety, Chardonnay, across five other grape varieties with varying FD symptoms expressions. Of the two largest test datasets, the true positive rate for Chardonnay reaches 98.48 % whereas for Ugni-Blanc it drops to 8.3 %, underlining the need for a multi-varietal training dataset to capture the diversity of FD symptoms. To obtain more transparent results and to better understand the model's sensitivity, we investigate its behavior using two visualization techniques, Guided Gradient-weighted Class Activation Mapping (GG-CAM) and the Uniform Manifold Approximation and Projection (UMAP). Such techniques lead to a more comprehensive analysis with greater reliability, which is essential for in-field applications, and more broadly, for all applications impacting humans and the environment.

4.3.1 Introduction

Flavescence dorée (FD) is a grapevine disease raising serious concern in Europe. This disease is caused by several phytoplasmas classified according to their ribosomal DNA (16SrV subgroup C and D) (Filippin *et al.*, 2009) and grouped under the temporary name of *Candidatus Phytoplasma vitis* (Firrao *et al.*, 2004). The transmission of those phytoplasmas is mediated by infected leafhopper *Scaphoideus titanus* which transmit the disease when feeding on vine leaves. This leafhopper, native to North America, was first observed in France in 1958 (Bonfils et Schvester, 1960). Adult *S. titanus* have a limited dispersal distance, reaching 25m to 30m, although winds can cause passive dispersal over larger areas. However, the long distance spread is mainly caused by human activities, such as the trade of plant infested material (Chuche et Thiery, 2014). Despite a quarantine status and mandatory monitoring practices, FD is still spreading in Europe (Fanjul, 2017). FD symptoms appear gradually over the summer with a peak of expression at the end of August/September. Contaminated plants show a drooping appearance caused by a lack of lignification. Infected leaves roll up and change color : yellowing for white grape varieties and reddening for red grape varieties. FD causes the death of inflorescences and of berries (Chuche et Thiery, 2014). Symptoms expressed by

an FD-infected plant are the same as those expressed by *Bois noir*, another disease caused by phytoplasmas. Both diseases are gathered under the term of "grapevine yellows symptoms". To distinguish between them, laboratory analyses are necessary (Chuche et Thiery, 2014). Over the years, FD-contaminated plants will die or become less and less productive. In any case, they constitute an eventual source of contamination for the surrounding vineyards. In France, the uprooting of contaminated grapevines is mandatory. When 20 % of the grapevines in a parcel are contaminated, the whole parcel must be uprooted (Ministère français de l'agriculture, 2013). Winegrowers are thus exposed to significant economic losses. Since there is no control method for phytoplasma and no treatment to cure an infected plant, the disease management has focused on the FD vector : the leafhopper. In some at-risk areas in France, Italy and Switzerland, insecticide treatments are mandatory to limit the number of leafhoppers (Chuche et Thiery, 2014).

Another strategy complementary to vector control has been investigated by several research teams : the automatic detection of FD symptoms in the field. This approach is based on the idea that by quickly locating all the grapevines that might be contaminated with FD, winegrowers are able to eliminate all of the at-risk plant material. Winegrowers already scout their fields, but due to a lack of time or of manpower, this tedious task is not always carried out systematically on the entire vineyard. With a camera on a drone, a tractor or a ground robot, a complete search could be carried out in a relatively short time. The winegrower or agronomist would then only have to check the suspect grapevines spotted by the detection tool and take appropriate actions. To develop such an automatic detection tool, AL-Saddik *et al.* (2017) opted for non-imaging and proximal hyperspectral analyses to find the optimal wavelengths to identify FD. Using several spectral measurements and considering grape variety and symptom intensity, they developed specific vegetation indices for FD detection. These indices resulted in classification accuracies of more than 90 % on leaf-scale spectral signatures. Albetis *et al.* (2018) used multispectral drone images (combining Red-Green-Blue, or RGB, and Near-Infrared, or NIR) acquired on seven red grape varieties. From these images, 24 variables such as vegetation indices and biophysical parameters were computed and used to differentiate two study cases : 1) FD and wood diseases (Esca and Black Dead Arm) from asymptomatic grapevines, and 2) FD from wood diseases. Results were promising in the first study case, but more mixed in the second.

The processing of in-field images highlighted another challenge : the management of soil and shadow mixels, which are significant sources of misclassification. Cruz *et al.* (2019) used Convolutional Neural Networks (CNNs) on proximal RGB images to detect grapevine yellowing symptoms on a red grape variety, with particular emphasis on differentiating them from leafroll and *Stictiocephala biosonia* symptoms. Their model achieved 98.96 % sensitivity and

99.40 % specificity. However, the images they used were collected under controlled conditions, which would not be suitable for an in-field automated detection tool. Nevertheless, numerous studies have shown the relevance of using CNNs to analyze in-field images on various crops (DeChant *et al.*, 2017; Fuentes *et al.*, 2018; Boulent *et al.*, 2019b). Despite the difficulties in analyzing field images due to complex background, leaf entanglements and shadows, CNNs are robust enough to overcome those issues and find discriminating features to identify several diseases on different crops.

CNNs provide a modeling approach that is part of deep learning methods. CNNs consist of neural networks that can be trained to fit their internal parameters for a specific task based on a large amount of training data. More specifically, CNNs are primarily a set of convolutional and non-linear layers that extract hierarchical representations to solve a task defined through a loss function. For more detailed information on deep learning and CNNs, please refer to LeCun *et al.* (2015) and Goodfellow *et al.* (2016). The deployment of CNNs has led to great progress in several vision-related applications, particularly in image content classification.

Although FD is a complex disease whose symptoms are not easy to detect, we hypothesize that CNNs can tackle this challenging task. Several confusion factors can be expected : 1) other phytosanitary problems, such as downy mildew or deficiencies, can be expressed through similar symptoms, especially on white grape varieties (e.g. yellowing) ; 2) a combination of health issues can occur on the same leaves, particularly at the end of the season, which complicates the visual identification of FD ; 3) the symptoms of FD can be expressed differently between varieties beyond the main two grape families (red and white) through variations in curling or coloring for example (AL-Saddik *et al.*, 2017; Albetis *et al.*, 2018) ; and 4) FD symptoms also seem to fluctuate from one year to the next, even for the same grape variety (Albetis *et al.*, 2018).

The objective of this study is to achieve the automatic pre-identification of FD symptoms in several white grape varieties from images taken in the field. The term "pre-identification" refers to the fact that a laboratory analysis must be performed to ensure that a grapevine is indeed contaminated with FD. Our level of detection is downstream : the idea is to visually detect FD-like symptoms. Therefore, in this study, any mention of FD detection refers to this notion of visual "pre-identification". The multi-varietal dimension is studied here because, for professional use, an automatic detection tool would be of real interest only if it could be used on several grape varieties. Acquiring images of all the targeted grape varieties would be expensive and very challenging. However, the symptoms, even within the same color family, can be quite different. So, could a model trained on one grape variety be effective on other grape varieties ? To answer this question, we train a classification model on Chardonnay images and transform it into a Fully Convolutional Network (FCN, Shelhamer *et al.* (2017)). We then

test it on several independent datasets originating from different sources. To provide greater transparency to our model and its results, we use visualization techniques to understand which image features the model uses for its predictions. This also helps us understand the differences in precision obtained on the different grape varieties. By identifying the strengths and weaknesses of our model (along with ways to improve its robustness and reliability) we take one more step towards the development of a professional tool.

4.3.2 Material and Methods

Method overview

To provide greater clarity, this preliminary section presents an overview of the processing steps performed in this study (Figure 4.1). To develop a tool to automatically detect FD symptoms, we first trained a CNN-based classifier. As an output, this model indicates whether or not there are FD-like symptoms in an input image. As a way to help the model’s training, we ensure that it targets elements of interest in the grapevines by annotating portions of images (mostly leaves). In the FD class of the training dataset, there are only Chardonnay leaves. The training process follows a classical approach : a ResNet18 is pre-trained on ImageNet and then completely fine-tuned with the FD training dataset. In fact, as our training dataset is quite small, it is advantageous to use weights from a prior training on a large dataset to retrieve the learned features as a starting point for our model training (Tajbakhsh *et al.*, 2016; Tan *et al.*, 2018).

The image-level binary classifier output (e.g. a unique FD or Non FD predicted label) is not the most suitable for our application. Indeed, a rough localization of the affected regions in the analyzed images would be much more appropriate. It would offer a more complete analysis while targeting suspect areas, bringing more transparency to the results and thus meeting the needs of the users. To achieve this rough localization, we opted to transform our classifier into an FCN. This transformation is simple in practice and it does not require any retraining to generate coarse class segmentation maps during inference (Shelhamer *et al.*, 2017).

To test our model, we use several FD image datasets. Two are from personal acquisitions (*Pers.* datasets). These have the advantage of being quite large (153 images for the one on Chardonnay and 155 for the one on Ugni-Blanc) and they have a certain independence from the training dataset, as the images come from separate plots. However, they allow a possible bias : the images used for training and testing were taken during the same year, by the same team, and with similar cameras. To address this issue, we introduce a dataset from an external source (*Ext.* datasets), provided by FREDON (Fédération Régionale de Défense contre les Organismes Nuisibles) Aquitaine. Although small in size (34 images), it

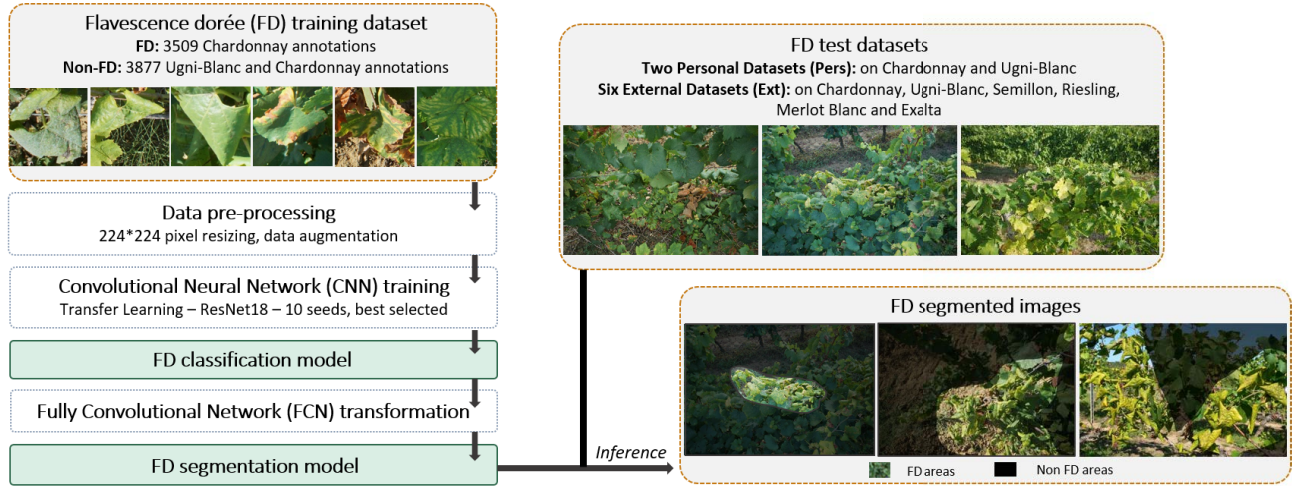


FIGURE 4.1 – Processing setup overview.

is of significant value, as it ensures that our model can work on images from completely independent sources. Furthermore, it also provides an opportunity to evaluate our model on several white grape varieties.

Datasets

Image collection process

To train and test the FD pre-identification model, we use RGB images captured with standard cameras. The training images were taken in 2018, in France, in the Cognac (45°41'24"N, 0°19'12"W) and Limoux (43°2'60"N, 2°12'36"E) regions. The data acquisition was carried out from mid-August to mid-September, allowing the recording of varied FD symptom intensities. During the scouting, we mainly found FD-infected grapevines in Chardonnay plots, hence our constraint to focus on this grape variety for the FD class. We also encountered combinations of FD symptoms with symptoms of downy mildew, pest damages and phytosanitary treatment residues. Thus, even though our training dataset for the FD class is single-varietal, it contains some diversity in the symptoms' expression. During our scouting, we also encountered other plant health problems in Chardonnay and Ugni-Blanc plots : downy mildew, Esca, Black Dead Arm, damage caused by acarions, mineral deficiencies (in magnesium, potassium, manganese and iron) as well as burns due to pesticides. Two plots were selected as acquisition sites for the test datasets (referred to as *Pers.* datasets for "Personal" acquisition) : one of Ugni-Blanc and one of Chardonnay. Thus, the images used to train and test the model came from separate plots.

Image acquisition was split based on two approaches : some images were captured using a

hand-held camera, while others were captured from a camera mounted on a two-meter pole. The latter approach allowed a wider view of the grapevines. Due to the changing vineyard structure (inter-row distances, row heights, trimming type), it was not possible to define fixed acquisition parameters. The distance between the camera and the foliage as well as the acquisition angle thus vary from one photo to another. Hence, the photographs have slightly different spatial focus and resolution. No restrictions in terms of weather conditions were established, and the only requirement for acquisition was that the leaves had to be dry. During the acquisitions, images were taken from both shaded and sunny rows, resulting in varying light exposures.

The set of 34 images prepared by a FREDON Aquitaine technician were reserved for testing¹. This set covers six grape varieties, with 3 to 12 images per variety : Chardonnay, Ugni-Blanc, Merlot Blanc, Riesling, Exalta and Semillon. All the images contain symptomatic FD grapevines. Figure 4.2 shows the typical expressions of the different grape varieties in this external (*Ext.*) dataset. Here, Chardonnay expresses a significant rolling with yellowing of a shade that differs from that of the training set. Exalta has rolled leaves with a withered aspect, sometimes with yellowing veins. Similar coloring is found on some Semillon leaves. Merlot Blanc and Riesling leaves show medium curling and yellowing. As for the Ugni-Blanc, it has little curling but significant coloration, sometimes with yellowing of the veins. Such variations in expression suggest generalization difficulties for our Chardonnay model on grape varieties expressing little curling. Although small in size, this test dataset is particularly relevant for the evaluation of our model because it was acquired by researchers not involved in this study. We know neither the camera model nor the year of acquisition, and had no control over the acquisition conditions. It therefore constitutes a realistic robustness test.

The Table 4.1 summarises the information about the datasets used to train the CNN classifier and to evaluate the FCN model.

Data preparation

Taxonomy definition and annotation A binary taxonomy was chosen for our classification model : FD versus Non FD. This choice was made primarily because it simplifies the annotation step. While several diseases were present in the surveyed plots, and these sometimes appeared in combinations, we do not have the expertise required to identify all the diseases we observed. The binary taxonomy therefore allows us to put everything that does not look like FD into a generic "Non FD" class. This class contains healthy grapevines, but also includes weeds, soil, sky and trellising elements. On the other hand, the FD class only contains leaves with FD-like symptoms, although no laboratory tests were conducted to

1. Available online : <https://flavescencedoree.jimdofree.com/>

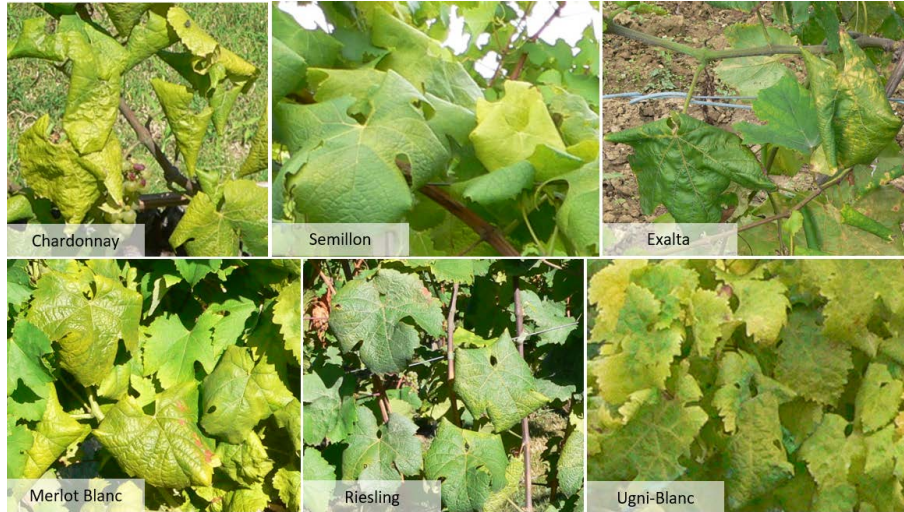


FIGURE 4.2 – Illustration of Flavescence dorée symptoms for the different grape varieties found in the *Ext.* test dataset. Source : D. Vergnes, FREDON Aquitaine.

confirm that FD was truly present.

To train the classifier, we manually extracted diverse samples from the captured images. For the plant elements, a similar unit was used throughout the annotation : the leaf or the bunch of grapes. We identified the minimal enclosing square for each image region of interest. Around 7,300 annotations were generated to train the classifier, with 3,509 for the FD class and 3,877 for the Non-FD class. Some samples within these classes are shown in Figure 4.3. Before each training session, the samples are randomly separated into training samples (85 %, or 6 279) and validation samples (15 %, or 1 107). In order to account for this random split in our evaluation results, we run a series of 10 independent experiments while shuffling these sample sets.

To test the classification model post-FCN conversion, we manually generate segmentation maps by coarsely delimiting the image regions potentially contaminated by FD. Segmentation maps were also manually generated on 15 images from the training set to help choose the best model out of our 10 experiments.

Image preprocessing The samples used for training our binary classifier are resized to 224×224 pixels. Since we use transfer learning based on ImageNet weights, we normalize the images using the mean and standard deviation values given for the RGB channels on the PyTorch website². To increase the model’s robustness to geometric and dynamic effects and to add diversity to our images, the following data augmentation operations are applied on the

2. https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html

TABLE 4.1 – Summary of datasets used to train the CNN classifier and to evaluate the FCN model. *As we extracted samples for both classes in some images, we only indicate the total images’ count for the CNN Training dataset.

Dataset Name	Grapevine Varieties	Number of images
CNN Training Dataset	FD class : Chardonnay Non FD class : Chardonnay, Ugni-Blanc	Total : 894*
FCN Evaluation Datasets		
<i>Pers. Chardonnay</i> Test	Chardonnay	FD class : 66 Non FD class : 87 Total : 153
<i>Ext. Chardonnay</i> Test	Chardonnay	FD class : 5 Non FD class : None Total : 5
<i>Pers. Ugni-Blanc</i> Test	Ugni-Blanc	FD class : 12 Non FD class : 143 Total : 155
<i>Ext. Other Grapevine Varieties</i> Test	Ugni-Blanc, Exalta, Merlot Blanc, Semillon, Riesling	FD class : 29 Non FD class : None Total : 29

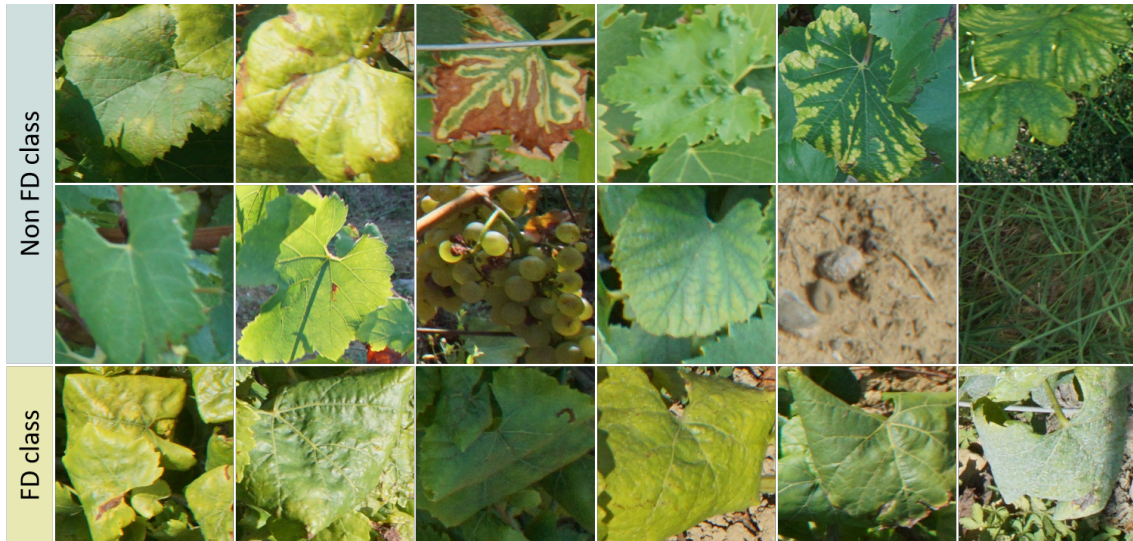


FIGURE 4.3 – Examples of samples from the CNN training dataset for FD and Non-FD classes.

images during training : image rotations (probability : 0.6), distortions (probability : 0.6), flips (probability : 0.8) and changes in brightness (probability : 0.6). These transformations are applied randomly to each sample using Augmentor³. For FCN inference, only the normalization was reapplied.

Methods

Training

As mentioned earlier, the FD pre-identification model is based on a fine-tuned ResNet-18 architecture (He *et al.*, 2016) pre-trained on ImageNet. This architecture is well-known and performed well in our previous experiments despite its simple nature (Boulent *et al.*, 2019a). The hyperparameter values are fixed based on several trials that involved the learning rate definition method proposed by Smith (2018) implemented in the FastAi library⁴. The values of the hyperparameters are provided in Table 4.2. The framework used to fine-tune and test the model is available on GitHub⁵.

TABLE 4.2 – Hyperparameters used to trained the ResNet-18 classifier.

Hyperparameter Name	Hyperparameter Value
Learning Rate	1e-04
Learning Rate Scheduler	Step size : 5, Gamma : 0.7
Optimizer	Adam
Loss Function	Cross Entropy
Batch Size	64
Epochs	15

Inference

Once trained, our image classifiers were converted to FCN models for semantic segmentation following the methodology detailed by (Shelhamer *et al.*, 2017). By replacing the fully connected layer with a 1x1 convolutional layer, a coarse segmentation map can be produced as the result of inference. Such limited resolution is caused by the use of pooling layers as well as the stride and padding values of the convolution layers. As a result, the output segmentation map is much smaller than the original image. To overlay with the original image, we resize this result to the input image’s size using bilinear interpolation.

3. <https://augmentor.readthedocs.io/en/master/>

4. <https://docs.fast.ai/>

5. <https://github.com/plstcharles/thelper>

We select the best CNN model out of our 10 experiments, using both validation accuracy and a visual quality assessment of the output segmentation maps on a handful of samples. The selected model is then evaluated on the four test sets : the two Chardonnay (*Pers.* and *Ext.*) as well as two other white grape varieties (*Pers.* and *Ext.*). Three metrics are used for performance evaluation : the True Positive Rate (TPR, Eq. 4.1) and the False Positive Rate (FPR, Eq. 4.2) to evaluate the quantity of both the correct and the incorrect detections generated by the model, and the Intersection over Union (IoU, Eq. 5.4), which provides a balanced overall look at the quality of the segmentation maps produced by the model. A Positive (P) sample is qualified as a "True" Positive (TP) if there was an intersection between the prediction and the annotation for the FD areas. A Negative (N) sample is qualified as a "True" Negative (TN) if the entire image was correctly predicted as Non-FD. The TPR was therefore more restrictive than a simple classification : it ensured that the detection was due to an FD area. Since for FD pre-identification, a false negative is more costly than a false positive, the decision threshold for the FD class has been set at 40 % instead of the 50 % usually used in binary classification.

$$TPR = \frac{TP}{P} \times 100, \quad (4.1)$$

where TP is the number of True Positive images and P is the number of Positive images ;

$$FPR = \frac{FP}{N} \times 100, \quad (4.2)$$

where FP is the number of False Positives images and N is the total number Negatives of images ; and

$$IoU_{AB} = \frac{A \cap B}{A \cup B}, \quad (4.3)$$

where A is the set of all FD pixels predicted by the model and B is the set of all pixels labeled as FD.

4.3.3 Results

During the fine tuning, all models converged quickly (see Figure 4.4). By the end of the first epoch, the accuracies ranged from 96.2 % to 98.7 % on the validation dataset for all 10 runs. From the 11th epoch, accuracy values became very close, with a difference between the minimum and maximum accuracies of only 0.59 % on average for the last 4 epochs. We transformed several of the obtained models with the highest precision values into FCNs, and used our validation dataset of 15 images to evaluate them. We decided to select the model

with the highest precision value at the 9th epoch.

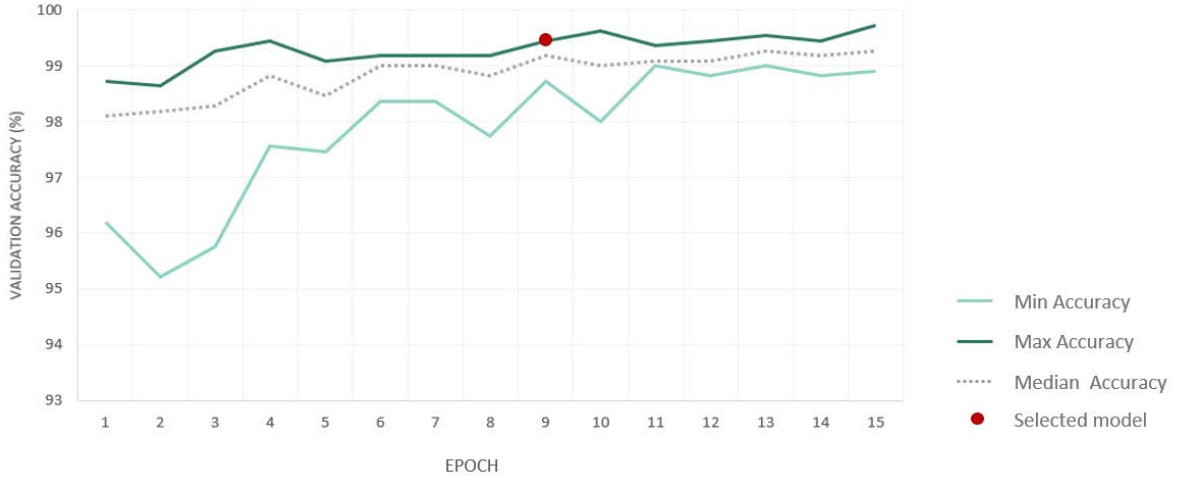


FIGURE 4.4 – Summary of accuracy values obtained on the validation dataset for the 10 runs.

Inference on Chardonnay datasets

For both the Chardonnay test datasets, the TPR values are very high : 98.5 % for the *Pers. Chardonnay* dataset (i.e. 65 TP for 66 P), 100 % for *Ext. Chardonnay* dataset (i.e. 5 TP for 5 P). There is a low number of false alarms for the *Pers. Chardonnay* dataset, with an FPR of 1.15 % (i.e. 1 FP for 87 N). These results highlight the detection power of the trained model. Next, the IoU values provide insight on the quality of the segmentation maps. The average IoU value is 0.53 for the *Pers. Chardonnay* dataset (with values ranging from 0.14 to 0.75), and 0.39 for the *Ext. Chardonnay* dataset (with values ranging from 0.26 to 0.6) (Table 4.3). Such values indicate a rather coarse segmentation quality. This was as expected, due to the simple classifier’s conversion to an FCN and its lack of advanced upsampling.

Figure 4.5 presents a set of typical predictions with minimum and maximum IoU values for the two Chardonnay datasets. The minimum IoU prediction on the *Pers. Chardonnay* dataset shows both over- and under-segmentation. However, two of the three leaves with FD symptoms are detected. Overall, we observe a better detection of FD when it is expressed on several leaves or on a branch. When symptoms affect only a few isolated leaves, there is a greater risk of non-detection (4.6, A and B). Leaves with only the underside visible are also under-detected (Figure 4.6, B). Furthermore, even if FD symptoms combined with downy mildew or residues of phytosanitary treatments are mostly detected, we noted a few cases of non-detection in this situation (Figure 4.6, A, B). Confusions with early downy mildew, with an embossed appearance, are present. With the minimum IoU prediction on the *Ext. Chardonnay* dataset,

TABLE 4.3 – Summary of inference results on both Chardonnay datasets, presenting the True Positive Rate (TPR), False Positive Rate (FPR), mean, min and max values of Intersection over Union (IoU). Acronyms : TP : True Positive, P : Positives, FP : False Positive, N : Negatives. * IoU min, considering only the True Positive predictions.

Test Dataset Name	Number of images	TPR	FPR	IoU mean	IoU min*	IoU max
<i>Pers. Chardonnay</i>	FD class : 66 Non FD class : 87 Total : 153	98.48 % 65 TP / 66 P	1.15 % 1 FP / 87 N	0.53	0.14	0.75
<i>Ext. Chardonnay</i>	FD class : 5 Non FD class : None Total : 5	100 % 5 TP / 5 P	N/A	0.39	0.26	0.6

another type of misdetection is related to the background. Grapevines, corn or a hedge in the background can sometimes be partly identified as FD. We also observed a case where another grapevine contaminated with FD in the background was detected. We did not anticipate these two situations during our acquisitions, as we took photos that mainly focused on one grapevine. This image, as well as the maximum IoU prediction on the *Ext. Chardonnay* dataset, also highlight the coarseness of the detection – the FD is properly detected, but the detection mask also includes some of the surrounding ground. Several detections were found, including soil or areas not contaminated with FD, but close to contaminated areas (Figure 4.6, C, E). Finally, false detections that are visually difficult to explain were also found (Figure 4.6, D).

As a result, despite the coarseness of the detections and some cases of confusion, the classification model converted to an FCN for large-scale inference is able to detect symptoms of FD on unseen images of Chardonnay grapevines. While the *Ext. Chardonnay* dataset is small, the metrics show that the model is able to generalize on images of the same grape variety, even if they come from another source. The results are very encouraging, as the expression of FD symptoms is slightly different between the two datasets ; the leaves are yellower in the *Ext. Chardonnay* images, but they remain quite green in the *Pers. Chardonnay* images. These two expressions, both present in the training set, are well-managed by the model.

Inference on several white grapevine varieties

Inference using the Chardonnay model on five other white grape varieties gave mixed results (Table 4.4). For the two Ugni-Blanc datasets, the Chardonnay model is unable to detect FD symptoms : 0 % of TPR for *Ext. Ugni-Blanc* dataset and 8.3 % of TPR for *Pers. Ugni-Blanc* dataset, with a very poor quality segmentation map for the only TP obtained (IoU value of 0.07). Even when leaves express curling, they are sometimes identified as Non-FD (Figure 4.7,

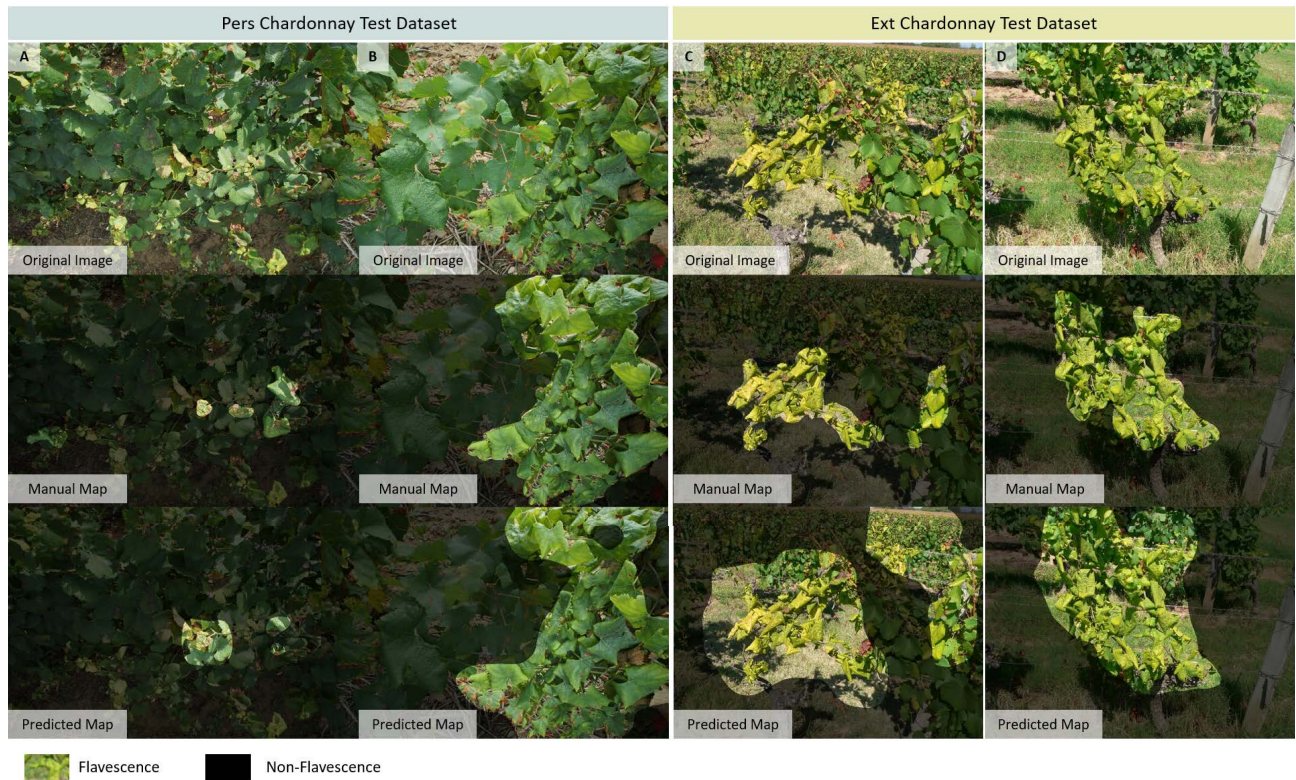


FIGURE 4.5 – Segmentation maps with minimum and maximum Intersection over Union (IoU) values on Chardonnay datasets. On the *Pers. Chardonnay* dataset, Images (A) with the lower IoU value, (B) with the higher IoU value. On the *Ext. Chardonnay* dataset, (C) with the lower IoU value, (D) with the higher IoU value.

C). The same phenomenon was observed on other grape varieties, such as Semillon (Figure 4.8, B) or Exalta (Figure 4.7, F), which was unexpected given that the model detects Chardonnay symptoms of FD characterized by rolling and significant yellowing (Figure 4.5, C and D). However, one result can be considered positive : the FPR of 2.87 % obtained on the *Pers. Ugni-Blanc* dataset is of the same order as the FPR value obtained on the *Pers. Chardonnay* dataset.

The inability of the Chardonnay model to identify FD on Ugni-Blanc could be caused by confusion with downy mildew or nutrients deficiencies. This hypothesis is based on the observation of the predictions for Exalta. In this grape variety, there were more detections on green leaves with heavy rolling. Yellow vein symptoms were classified as Non-FD (Figure 4.7, A and B). This symptom's expression has not been observed on the visited Chardonnay plots and is therefore absent from the training dataset. For Exalta and Semillon varieties, even if the TPR is high, the segmentation is of poor quality, with many under-detections (Figure 4.8, A and B).

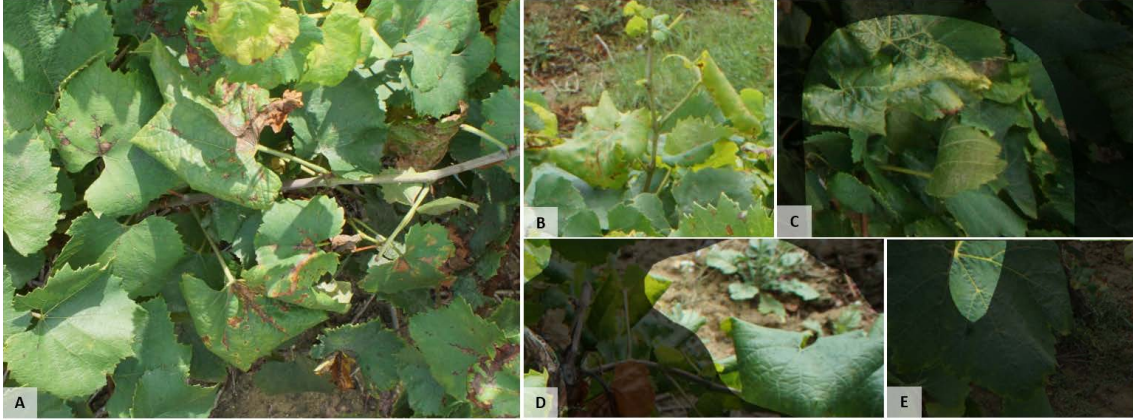


FIGURE 4.6 – Examples of false predictions from the *Pers. Chardonnay* Test dataset. (A), (B) Undetected Flavescence dorée (FD) symptoms, (C), (D) Over-detection close to FD symptomatic areas, E : Visually unexplainable false FD detection.

TABLE 4.4 – Summary of inference results on other grapevine varieties’ datasets, presenting the True Positive Rate (TPR), False Positive Rate (FPR), mean, min and max values of Intersection over Union (IoU). Acronyms : TP : True Positive, P : Positives, FP : False Positive, and N : Negatives. * IoU min, considering only the True Positive predictions.

Test Dataset Name	Number of images	TPR	FPR	IoU mean	IoU min*	IoU max
<i>Pers. Ugni-Blanc</i>	FD class : 12 Non FD class : 143 Total : 155	8.3 % 1 TP / 12 P	2.87 % 4 FP / 139 N	N/A IoU for the only TP : 0.07		
<i>Ext. Ugni-Blanc</i>	FD class : 5 Non FD class : None Total : 5	0 % 0 TP / 5 P	N/A	N/A	N/A	N/A
<i>Ext. Exalta</i>	FD class : 12 Non FD class : None Total : 12	83.3 % 10 TP / 12 P	N/A	0.13	0.02	0.31
<i>Ext. Merlot Blanc</i>	FD class : 3 Non FD class : None Total : 3	100 % 3 TP / 3 P	N/A	0.18	0.01	0.46
<i>Ext. Riesling</i>	FD class : 4 Non FD class : None Total : 4	100 % 4 TP / 4 P	N/A	0.27	0.17	0.39
<i>Ext. Semillon</i>	FD class : 5 Non FD class : None Total : 5	100 % 5 TP / 5 P	N/A	0.14	0.01	0.24

On Riesling and Merlot Blanc varieties, the segmentation is of relatively better quality, with average IoUs of 0.27 and 0.18, respectively. Again, cases of under-detection were noted (Figure 4.8, C). One possible explanation could be the difference in resolution between the images used

for training and those used for inference. Indeed, in images found in Figure 4.8 C and D, the same leaves were photographed with a more or less close-up view. On the close-up view, FD is detected, while on the farther view, symptomatic leaves are labelled as Non-FD. Furthermore, as observed in the Chardonnay datasets, there is an over-detection in the predictions around areas identified as symptomatic (Figure 4.7, D), but there is also over-detection related to the background that is barely interpretable (Figure 4.7, E).

In summary, the Chardonnay model does not work to detect FD symptoms on the two Ugni-Blanc datasets. However, the low FPR obtained on the *Pers. Ugni-Blanc* dataset may translate to significant robustness in a real-world application. For the other grape varieties, even if the TPR is high, the average IoUs are very low. In particular, many under-detections were noted. Therefore, the segmentation maps do not have the required quality for an automatic detection tool.



FIGURE 4.7 – Examples of false predictions from *Ext.* datasets. (A) Original image of Exalta leaf with Flavescence dorée (FD) symptoms; (B) Prediction associated to image A, only the rolled areas of the leaf are identified as FD; (C) Undetected FD symptoms on Ugni-Blanc grapevine variety; (D) Over-detection close to FD symptomatic areas; (E) False detection on non-grapevine elements; and (F) Undetected FD symptoms on Exalta grapevine variety.

4.3.4 Discussion

Providing a model with high accuracy for FD symptoms recognition is a first step towards the development of an automatic detection tool, but is not sufficient in itself. Indeed, in agriculture, as in all other fields of application that may have social and environmental impacts, it is necessary to move towards artificial intelligence tools whose results and underlying processes are transparent, understandable and explainable. Relying on opaque results does not provide

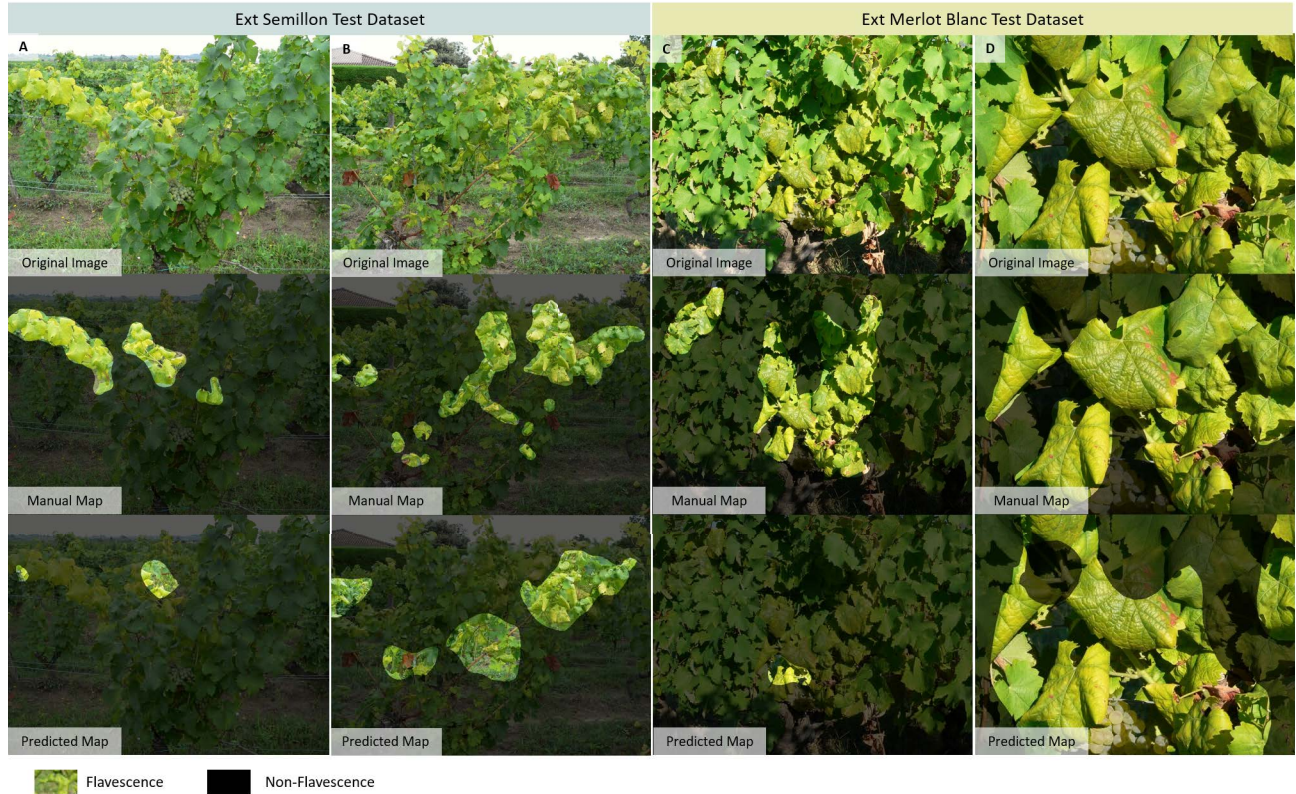


FIGURE 4.8 – Segmentation maps with minimum and maximum Intersection over Union (IoU) values on two *Ext.* datasets. On the *Ext. Semillon* dataset : Images (A) with the lower IoU value, and (B) with the higher IoU value. On the *Ext. Merlot Blanc* dataset : (C) with the lower IoU value, and (D) with the higher IoU value.

a sufficient degree of reliability for real-world applications. Wrong predictions in the field could have severe economic consequences for farmers. Without transparency, it would be difficult for farmers to have confidence in new AI-driven tools. Explainable AI refers to all the methods and techniques put in place to allow humans to understand the decisions and results produced by AI systems, thus avoiding a black box effect (Doran *et al.*, 2018; Holzinger *et al.*, 2017; Arrieta *et al.*, 2020).

Here, we take a step towards greater transparency by increasing the interpretability of our system regarding two aspects : 1) the multi-varietal application capability of the model trained only on Chardonnay FD symptoms, and 2) the general robustness of the model. For this purpose, we use two visualization techniques. The first one, Guided Gradient-weighted Class Activation Mapping (GG-CAM, Selvaraju *et al.* (2019))⁶, provides a finely detailed and discriminative view of class activations. This allows us to observe which characteristics of an image are used to predict a targeted class. The second visualization technique, Uniform

6. Python library used : Captum (Kokhlikyan *et al.*, 2019)

Manifold Approximation and Projection (UMAP, McInnes *et al.* (2018))⁷, is a dimensionality reduction algorithm that preserves the global structure of the projected data. This allows us to visualize the spatial arrangement of the image embeddings inferred by the model. Here, we use UMAP to observe the closeness of the embeddings according to their class and grape variety.

To improve the readability of this analysis, we used a subset of our evaluation datasets to create several visualizations. This subset contains 360 patches extracted from the images within the six evaluation datasets. Specifically, 30 patches were extracted for each of the following 12 categories : *Pers. Chardonnay*, Healthy ; *Pers. Chardonnay*, Downy Mildew ; *Pers. Chardonnay*, Deficiencies ; *Pers. Chardonnay*, FD ; *Ext. Chardonnay*, FD ; *Pers. Ugni-Blanc*, Healthy ; *Pers. Ugni-Blanc*, FD ; *Ext. Ugni-Blanc*, FD ; *Ext. Exalta*, FD ; *Ext. Merlot Blanc*, FD ; *Ext. Riesling*, FD ; *Ext. Semillon*, FD.

Multi-varietal application

We expect that the need for FD detection will appear for many grape varieties. If images have to be collected on all the targeted grape varieties, the task would become increasingly challenging and expensive. Our model, trained on a single symptomatic white grape variety, resulted in mixed performances when tested on other varieties. By applying GG-CAM to our images, the characteristics typically used by the model to associate FD and Non-FD classes are highlighted (Figure 4.9). For both classes, leaf morphology seems to be important. For the Non-FD class, it is mainly the leaf teeth that stand out (Figure 4.9 A, B, C, G), but the orientation of the veins also appears to be relevant (Figure 4.9 A). For the FD class, morphologically, it is mainly the curl, the sinus shape and the vein orientation that emerge (Figure 4.9 D, E, F, H). Leaf texture also seems to be a discriminating feature, as an embossed, scaly aspect of the leaves appears (Figure 4.9 E, F, H). Color is also a decisive feature. For the FD class there is a yellow dominance, whereas for the Non-FD class green and purple spots are frequently seen. Canes and petioles are mostly correctly associated with the Non-FD class (Figure 4.9 A, F, H). The model’s discriminative leaf features spotted through GG-Cam are of interest since they reflect the characteristics used by experts in the field for their diagnosis. Such general class feature descriptions provide a better understanding about FD false negative predictions across the different grape varieties. The Ugni-Blanc, which here expresses little curling, seems to be associated with the Non-FD class because of the toothed aspect of the leaf, which remains despite FD contamination (Figure 4.9 G). In some cases, including those showing partial curling, both classes have strong activations (Figure 4.9 F, H), but the Non-FD class is predicted by the model. This problem may have been accentuated here by the size of

7. The Python library used here : UMAP (McInnes *et al.*, 2018)

the image patch as well as by the core size of the average pooling layer. Some FD symptoms do not appear in the GG-CAM maps, including vein discoloration and FD-symptomatic leaves without curling.

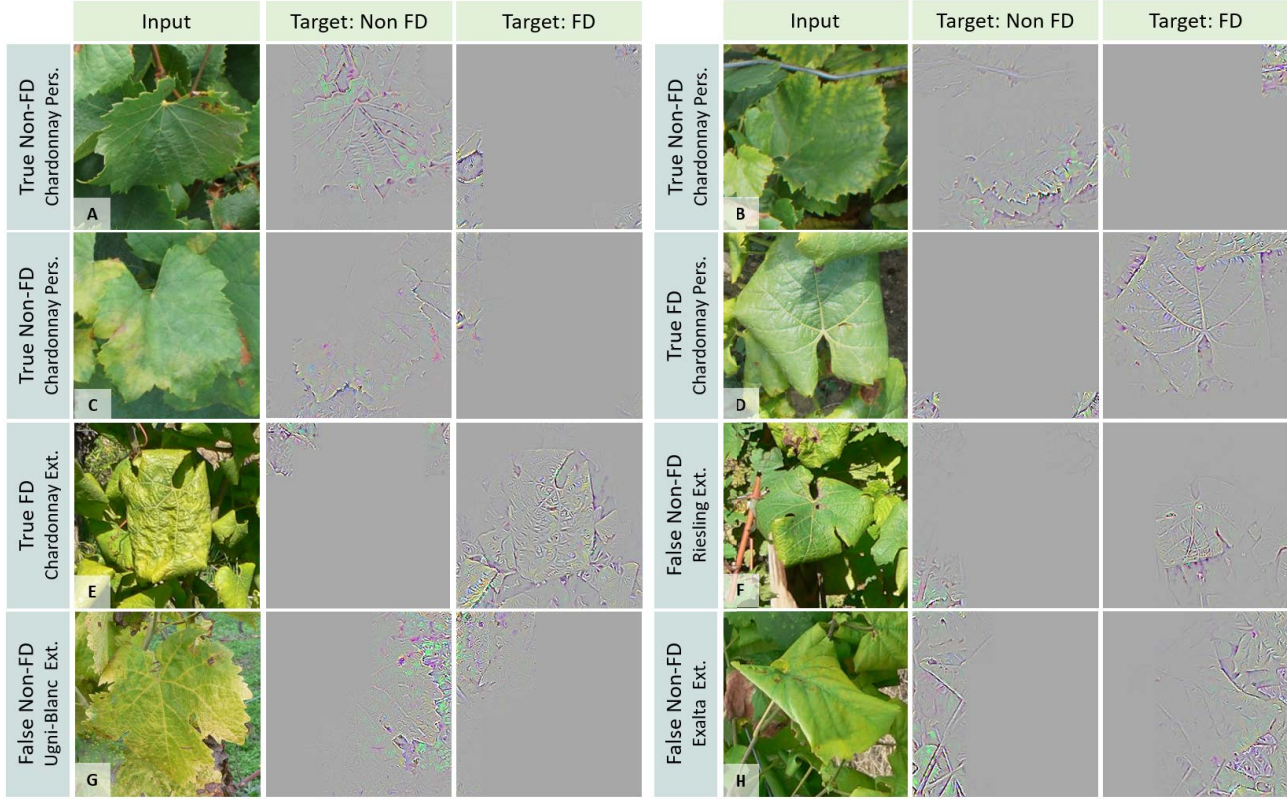


FIGURE 4.9 – Illustration of representative features from Guided Grad-CAM that are used by the model to predict the Flavescence dorée (FD) and Non-FD classes. A - C : Non-FD images, D - H : FD images. The predicted label and the source dataset are given next to each image.

The unawareness of these symptoms, which are very common for some grape varieties, is also revealed through UMAP visualization (Figure 4.10). Indeed, in this visualization, two clusters are found at each end : the true predictions of Non-FD and FD. In the Non-FD cluster, healthy Ugni-Blanc and healthy Chardonnay share the same space. Downy mildew and deficiencies embeddings are also quite close. At one end of this cluster, we observe proximity to the true Non-FD positives that correspond to Chardonnay deficiencies and downy mildew patches with false Non-FD negatives of Ugni-Blanc.

In the true FD prediction cluster, all of the Chardonnay patches from the *Pers.* and *Ext.* datasets are found with the majority grouped at the 0.8 :1 end of the x-axis. Elsewhere in the cluster, many Semillon, some Exalta, Riesling, Merlot Blanc and, to a lesser extent, Ugni-Blanc can be found. In between our two main clusters of true predictions is a zone of

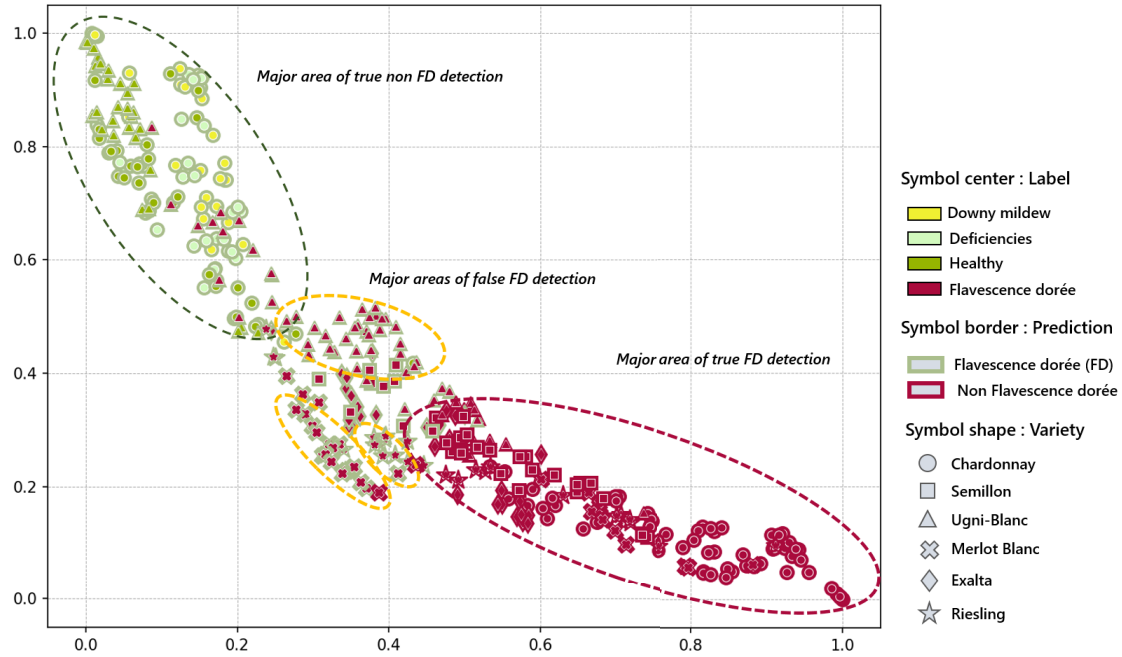


FIGURE 4.10 – Grape varieties in the model space : UMAP visualization of the embeddings based on an arbitrarily-set seed.

false negative FD embeddings. In this area, there are micro-clusters based on grape variety. These are certainly influenced by the closeness of the expressed symptoms among the same grape variety, but potentially also by having images from the same source. Despite this possible bias, the UMAP visualization confirms the hypothesis that the model fails to detect FD symptoms on Ugni-Blanc because of their closeness to deficiencies and to downy mildew symptoms. For false detections on the other grape varieties, a greater distance is observed from the cluster of good Non-FD predictions, which expresses the model’s ambiguity towards symptom expressions not directly seen during training. To improve this model and make it multi-varietal, new acquisitions would be necessary to include images with grapevines presenting symptoms with intermediate curling intensities and vein colorings.

Towards a robust detection model

Although the results obtained on Chardonnay are promising, further acquisitions are needed to increase our model’s reliability. Indeed, even for only FD on a given grape variety, symptoms may be expressed differently from one year to another. In addition, the general phytosanitary situation of a field can be very different from one year to another. Since FD is a late-season disease, grape leaves may already be damaged when FD symptoms appear. Acquisitions over several years will provide the needed data on the co-occurrence of phytosanitary problems.

The weaknesses of our model can also be targeted using the feature maps produced with GG-CAM (Figure 4.11). Indeed, some elements of the image stand out for the FD target when they should not, such as the trellising wire (Figure 4.11 A, D, E, F), the soil (Figure 4.11 A, D, F and Figure 4.9 H), some shadows or leaf overlays with triangular or rolled shapes (Figure 4.11 D and Figure 4.9 A, H), as well as, in rare cases, grapes, petioles or canes (Figure 4.11 A, B, C). To overcome this, targeted acquisitions to increase training image diversity could be conducted. Intervention at the model level could also be performed, e.g. by adding the annotation mask as an extra channel during training to act as an attention map, or with explanatory interactive learning, as in Schramowski *et al.* (2020) for plant phenotyping. These techniques would allow us to ensure that the predictions obtained are based on relevant features.

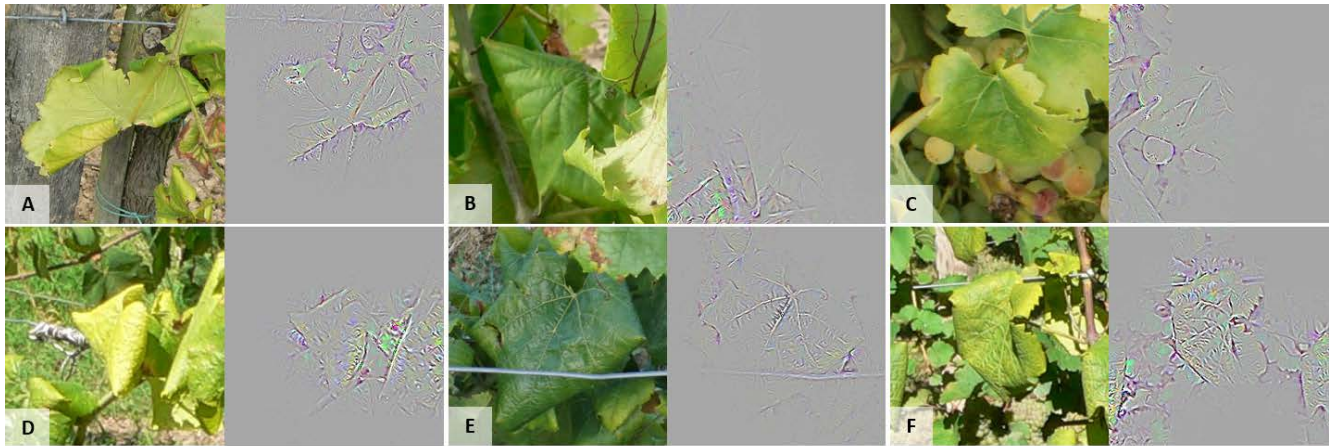


FIGURE 4.11 – Illustration of irrelevant features from GG-CAM that are used by the model to predict the Flavescence dorée (FD) class. A : Trellising wire, petioles, canes, soil and rolled shapes, B : Canes, C : Canes and grapes, D : Trellising wire and triangular shadow shape, E : Trellising wire and triangular shadow shape, and F : Trellising wire and soil.

4.3.5 Conclusion

In this study, we sought to automatically pre-identify FD symptoms on several white grape varieties. We trained a CNN model and turned it into an FCN so that it produced segmentation of areas expressing FD symptoms. In the training dataset, all symptomatic annotations were based on Chardonnay grapevines. During evaluation, this model obtained high accuracy on Chardonnay images, both on the personal and on the external datasets. Using the GG-CAM visualization technique, weaknesses in the model were identified, such as sensitivity to background or to trellising elements. Further targeted acquisitions and re-training could

solve this problem, which is highly encouraging. A larger external dataset obtained over several growing seasons would nevertheless be necessary to ensure that this model is indeed operational on Chardonnay grapevines.

For the other grape varieties, the results obtained were more nuanced, and showed how different symptoms' expressiveness could be problematic for FD pre-identification. As expected, the model performed poorly for the grape variety expressing the least amount of curling, i.e. Ugni-Blanc. We were able to confirm this hypothesis using GG-CAM visualizations. The embeddings extracted through convolutions for FD-contaminated Ugni-Blanc leaves are close to those of leaves affected by downy mildew and deficiencies. For the other grape varieties, the performance is mixed, with very high prediction rates. For a model trained with a grape variety expressing a lot of curling, good results were achieved on grape varieties expressing curling to a lesser extent. This underlines the model's relative flexibility and validates the idea that the development of a multi-varietal model is achievable.

Our next steps are to develop a comprehensive overview of the various FD symptoms and to include them in a multi-varietal dataset. Appropriate precautions will be undertaken regarding all other diseases that may express yellowing and curling. Capturing images from these misleading negatives would help to establish a more accurate boundary between FD and Non-FD images. Finally, going back and forth between model training and field testing will allow us to develop a robust detection tool. Using visualization techniques, as we have done in this study, will help to provide a more complete model analysis, and thus help to achieve the reliability needed for in-field applications.

4.3.6 Acknowledgments

Many thanks to all the people who enabled the dataset acquisition used in this study : Microdrones Canada and France, the Rémy Martin vineyards, the Chambre d'Agriculture de l'Aude, more specifically, Céline Forget, the Fédération Départementale de défense contre les Organismes Nuisibles (FEDON) de l'Aude, and Julien Sendrous in particular, as well as all the owners and wine growers who gave us access to their vineyards and/or guided us in the image acquisition process. Special thanks to Eve Laroche-Pinel for sharing this in-field data acquisition. We would also like to thank the Fédération Régionale de défense contre les Organismes Nuisibles (FREDON) of Aquitaine, especially Dominique Vergnès, who shared his photos collected on various white grape varieties, as well as his expertise in identifying flavescente dorée symptoms. Thanks also go to Maël Moreni for his expertise on the learning rate definition technique and to Simon Morvan for his advice on FD biology and epidemiology.

Funding

This research project was funded by the Centre de Géomatique du Québec (CGQ), Mitacs, the Ministère de l'Économie et de l'Innovation (MEI) of the province of Québec, and Microdrones Canada.

Chapitre 5

Les RNCs face à différentes conditions d'acquisition : Analyse d'images à la résolution spatiale dégradée et acquises par drone

5.1 Introduction

Plusieurs paramètres sont à définir lors de l’acquisition d’images pour une prospection automatique, principalement le type de caméra, la plateforme sur laquelle elle sera montée, mais aussi l’angle et la distance de prise de vue. Le choix de ces paramètres est amené à varier d’une prospection à l’autre, notamment en fonction du relief, de la superficie et de l’aménagement des parcelles, ou encore des moyens matériels et financiers des agriculteurs. Dans l’idéal, un modèle de reconnaissance de problèmes phytosanitaires devrait avoir une certaine robustesse à la variation des conditions d’acquisition, lui permettant de généraliser sur des images acquises, par exemple, avec différents capteurs ou paramètres d’acquisition (comme l’altitude ou l’angle de prise de vue). Cependant, même pour une thématique identique, de tels changements de conditions d’acquisition vont impacter les statistiques des images. Face à des images à la distribution différant trop fortement de celle des images d’entraînement, le modèle risque d’être incapable de généraliser. Ce phénomène peut être observé dans plusieurs travaux visant au développement de technologies de reconnaissance de problèmes phytosanitaires. En appliquant leur modèle de reconnaissance basé sur un RNC sur des images acquises dans des conditions différentes de celles d’entraînement, Mohanty *et al.* (2016) ont vu le taux de bonnes prédictions de leur classification passer de 99,35 % à 31 %. Ramcharan *et al.* (2019) ont quant à eux observé une diminution de la valeur du F1-score de 0,25 entre l’ensemble de test utilisé lors de la conception du modèle et un ensemble d’images prises en conditions réelles d’application. Ainsi, dans un scénario d’acquisition variant entre l’entraînement et l’inférence, une divergence des ensembles de données (*dataset shift*) risque d’émerger. Ce problème est récurrent en apprentissage automatique et une multitude d’éléments peuvent en être la source (Storkey, 2009), comme une évolution temporelle, une différence géographique ou encore un biais induit par l’acquisition ou l’annotation. Toutes ces possibles différences entre les ensembles d’entraînement et d’inférence ainsi que leur impact sur la capacité d’analyse et de généralisation des modèles peuvent constituer un frein au développement et à l’utilisation d’outils de prospection automatique. En effet, constituer un ensemble d’entraînement pour la reconnaissance de problèmes phytosanitaires est déjà une tâche complexe. Cela requiert un échantillonnage sur plusieurs saisons, variétés et conditions de cultures (Barbedo, 2016; Boulent *et al.*, 2019b). Ajouter plusieurs paramètres d’acquisition complexifie donc encore la tâche, surtout si l’on vise une flexibilité d’application sur différentes plateformes. En effet, un échantillonnage d’une telle envergure peut être ardu avec certaines d’entre elles, dont le drone qui nécessite notamment l’obtention d’autorisations préalables au vol (Transport Canada, 2020).

Dans ces travaux, nous souhaitons explorer la limite d’application d’un modèle basé sur les

RNCs face à des images acquises dans des conditions différentes de celles utilisées pour leur entraînement. Plus spécifiquement, nous nous intéressons à deux éléments : la résolution spatiale et la plateforme d’acquisition. La résolution spatiale est un paramètre incontournable en géomatique mais moins étudié en traitement d’image, où on utilise d’ailleurs plutôt les notions de dimension de l’image et d’échelle de l’objet. Les RNCs ne sont pas naturellement équivariants à un changement d’échelle et n’y possèdent donc pas de réponse prévisible (Goodfellow *et al.*, 2016; Touvron *et al.*, 2019). Si l’on souhaite développer la capacité d’analyse multi-échelle d’un RNC, il est nécessaire de le confronter à des images aux résolutions variées lors de l’apprentissage. Ce changement de résolution peut cependant être réalisé de manière artificielle, notamment comme une opération d’augmentation de données. Touvron *et al.* (2019) ont constaté que la différence de pré-traitements entre les images en entraînement et en inférence engendrait une divergence de la taille observée des objets lors de ces deux phases. De façon similaire, la méthodologie que nous avons suivie jusqu’alors implique un redimensionnement de chaque patch en 224×224 pixels – taille de la couche d’entrée du réseau – pour l’entraînement puis l’envoi sans modification de taille de l’image en inférence par segmentation. Pour nos images d’entraînement, cette opération entraîne dans plus de 95 % des cas une perte de résolution. Du fait de cette différence de traitement, la taille observée des objets est supérieure en inférence par rapport à l’entraînement. Ce redimensionnement induit-il une capacité des RNCs à analyser des images à plus faible résolution ? Dans cette étude, nous explorons l’impact de la modification de la dimension des images en inférence sur les performances des RNCs. Nous voulons évaluer si ces modèles (entraînés sur des images acquises par perche) ont été suffisamment confrontés en entraînement à des objets de dimension restreinte afin d’être utilisés sur des images à plus faible résolution, voire issues d’acquisition par drone. Nous évaluons également s’il est possible de rendre un modèle apte à analyser des images aux paramètres d’acquisition variables (en particulier à l’échelle variable), sans pour autant avoir à passer par tout le processus d’acquisition et d’annotation classique. Pour cela, nous avons constitué des variantes d’un jeu d’entraînement nous permettant de déterminer si des modifications simples et peu coûteuses peuvent améliorer la capacité d’analyse des RNCs sur des images à plus faible résolution et provenant d’acquisition par drone.

5.2 Matériel et méthodes

5.2.1 Description générale de la méthode

Pour une plus grande clarté, cette section préliminaire présente une description générale de la méthode utilisée (Figure 5.1). Comme dans le Chapitre 4, cette étude se penche sur la

reconnaissance des symptômes de FD. Pour rappel, cette maladie inquiète énormément les viticulteurs européens puisque, malgré des mesures de lutte obligatoires, la FD continue de se répandre. Comme il n'existe pas de traitement spécifique contre les phytoplasmes qui en sont à l'origine, la lutte se fait par une réduction de la population de l'insecte vecteur, la cicadelle *Scaphoideus titanus*, et par l'arrachage des vignes contaminées. Sans arrachage, les vignes contaminées constituent une source de diffusion de la maladie au sein de la parcelle et dans celles alentours (Chuche et Thiery, 2014). Un outil de reconnaissance automatique de la FD doit donc se baser sur un modèle sensible, une omission étant plus coûteuse qu'une fausse détection.

Dans cette étude, plusieurs RNCs sont entraînés à reconnaître des symptômes de FD. Les différents RNCs développés sont entraînés avec les mêmes hyperparamètres. La différence entre eux se situe dans les ensembles d'entraînement. Trois variantes (décrites dans la Section 5.2.2) sont utilisées afin d'évaluer si de simples modifications d'un ensemble d'entraînement peuvent conduire à une meilleure capacité d'analyse d'images à la résolution spatiale plus faible et provenant d'une autre plateforme d'acquisition, le drone.

En inférence, sur un jeu de test indépendant, les modèles ont pour tâche d'analyser des images de résolution spatiale différente. Le changement de résolution spatiale est simulé par un redimensionnement de l'image. Pour éviter des effets de repliement de spectre (*aliasing*) lors de la diminution des images, la méthode d'interpolation `INTER_AREA` de la librairie OpenCV¹ est appliquée. Cette méthode d'interpolation calcule la valeur des nouveaux pixels en se basant sur la valeur moyenne des pixels dans un voisinage donné, dont la taille est définie selon le facteur de réduction appliqué à l'image. Cette méthode est particulièrement conseillée lors des réductions d'images afin d'éviter qu'un effet de moiré impacte la qualité de l'image.

Dans un premier temps, afin d'évaluer les performances des modèles en fonction de la résolution, ceux-ci sont confrontés successivement en inférence au même ensemble d'images acquises par perche, mais en trois dimensions différentes : 1) *Dimension originale* : 3976×2652 pixels ; 2) *Dimension intermédiaire* : 1729×1153 pixels ; 3) *Dimension équivalente à une résolution d'image acquise par drone* : 857×571 pixels.

Ensuite, afin d'évaluer les performances des modèles sur des images venant d'une autre plateforme (et présentant à la fois une différence en termes de résolution spatiale et d'angle d'acquisition), des inférences sont réalisées sur des images acquises par drone. Deux tailles d'image sont utilisées : 1) *Dimension originale*, correspondant à un vol à 10 mètres d'altitude : 3976×2652 pixels ; 2) *Dimension équivalente à un vol à 25 mètres d'altitude*, une hauteur de vol plus proche de l'opérationnel : 1616×1078 pixels.

1. https://docs.opencv.org/3.4/da/d54/group__imgproc__transform.html

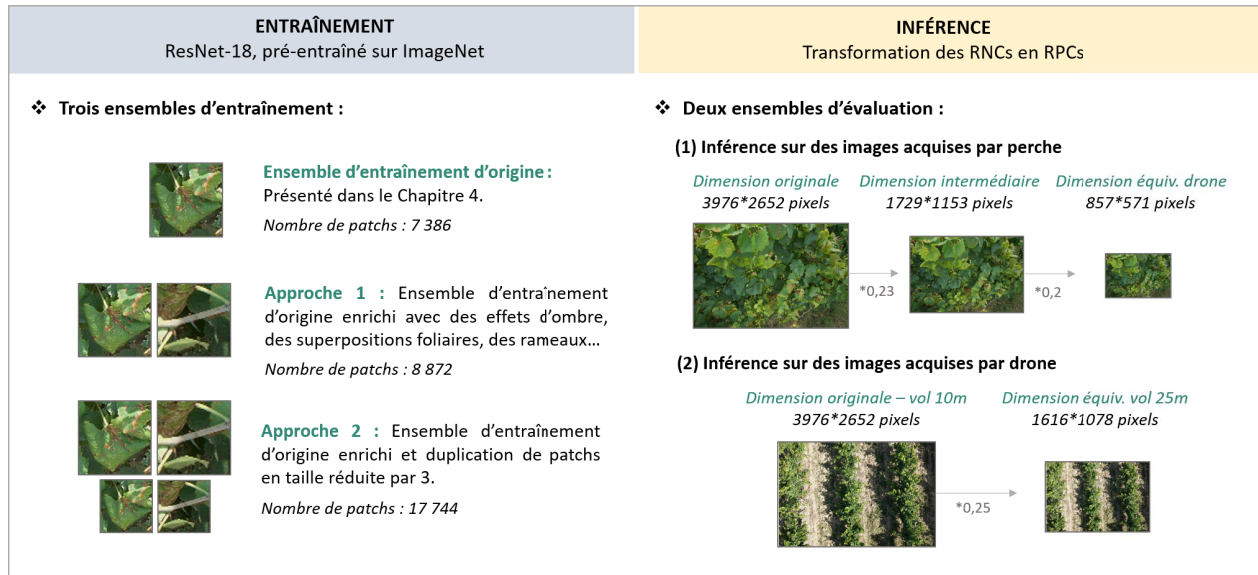


FIGURE 5.1 – Illustration des différentes approches d'entraînement et d'inférence adoptées.

5.2.2 Données

Ensembles d'entraînement

Les images utilisées pour entraîner les RNCs proviennent de l'acquisition précédemment détaillée dans le Chapitre 4, Section 4.3.2. Pour rappel, cette acquisition a eu lieu en France, en 2018, dans des vignobles aux alentours des villes de Cognac (45°41'24"N, 0°19'12"W) et de Limoux (43°2'60"N, 2°12'36"E). Elle a été réalisée dans des parcelles plantées en cépages blancs, plus précisément en Chardonnay ou Ugni-Blanc pour les parcelles asymptomatiques en FD, et en Chardonnay pour les parcelles symptomatiques en FD. Des photos Rouge-Vert-Bleu (RVB) ont été acquises au moyen d'un appareil photo tenu à la main ou monté sur une perche. Ces images ont ensuite été annotées, permettant de générer des patches associés soit à la classe "FD" ou à la classe "Non FD". Avant d'être envoyés au RNC, les patches sont redimensionnés en 224×224 pixels, incluant ainsi déjà une forme de dégradation de résolution spatiale. Ces images sont également normalisées en utilisant les valeurs de moyenne et d'écart-type données pour chaque bande sur le site internet de PyTorch². Les trois variantes du jeu d'entraînement sont présentées ci-après.

Ensemble d'origine Ce jeu d'entraînement contient 7 386 patches, 3 509 pour la classe "FD" et 3 877 pour la classe "Non FD". Une description complète de cet ensemble et des images qu'il contient peut être retrouvée dans le Chapitre 4, Section 4.3.2.

Ensemble de l'approche 1 Ce jeu d'entraînement regroupe les patches de l'ensemble d'origine et 1 586 annotations supplémentaires, pour un total de 8 872 patches (3 770 pour la classe "FD"

2. https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html

et 4 867 pour la classe "Non FD"). Ces annotations ont été réalisées sur les mêmes images que celles de l'ensemble d'origine. Ces nouvelles annotations ont un double objectif : 1) Pallier aux faiblesses du modèle développé au Chapitre 4 et repérées par l'analyse des mauvaises prédictions et des visualisations de GG-CAM (Selvaraju *et al.*, 2019) ; 2) Augmenter le nombre d'exemples de caractéristiques plus fréquemment présentes dans les images acquises par drone. Ainsi, ces nouvelles annotations concernent des zones ombragées, des mauvaises herbes, des grappes, des rameaux, des fils et poteaux de palissage, des effets de superposition foliaire (particulièrement ceux faisant apparaître une forme triangulaire), des feuilles contaminées par la FD mais qui conservent un aspect dentelé, des feuilles exprimant les premiers symptômes d'une contamination au mildiou ainsi que des feuilles exprimant à la fois de la FD et du mildiou.

Ensemble de l'approche 2 Ce jeu d'entraînement contient les patches de l'ensemble de l'approche 1 ainsi que leur duplication dans une taille réduite. Il regroupe donc 17 744 patches, 7 540 pour la classe "FD" et 9 734 pour la classe "Non FD". L'objectif de cette duplication en taille réduite est de confronter les RNCs à des images à plus faible résolution spatiale dès la phase d'entraînement. Les RNCs que nous avons entraînés tout au long de nos expérimentations sont déjà confrontés à cette perte de détails puisqu'un redimensionnement en 224×224 pixels a été appliqué systématiquement. La démarche de duplication vise à pousser cette dégradation plus loin en diminuant préalablement la taille des patches par un facteur 3, permettant ainsi d'obtenir plus de 50 % des patches dont la taille est inférieure à 224×224 pixels (Figure 5.2). Nous avons choisi de combiner dans cet ensemble à la fois les patches de taille originale et de taille réduite afin de favoriser le développement d'un modèle capable de gérer différents niveaux de résolution spatiale.

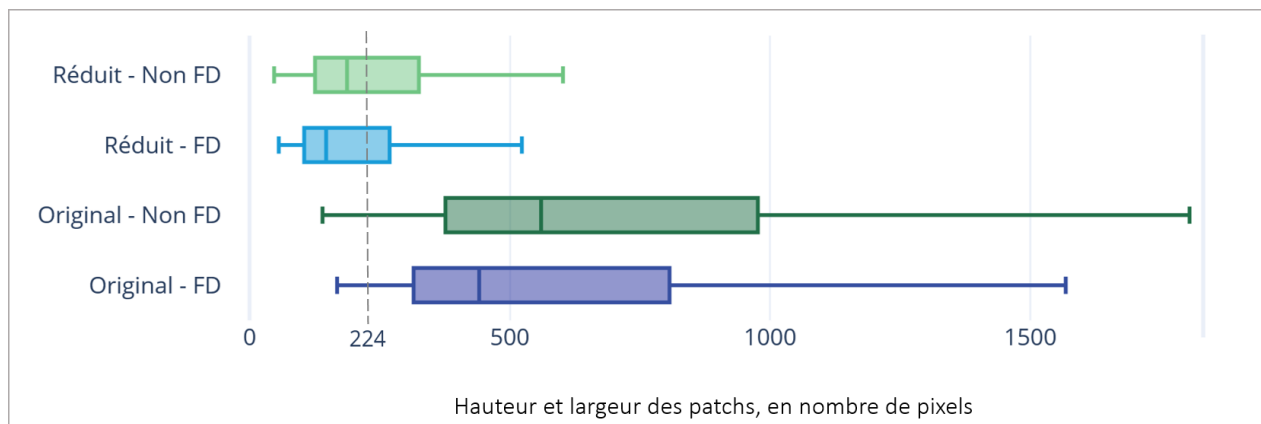


FIGURE 5.2 – Diagramme en boîte représentant la taille des patches en nombre de pixels pour chaque classe, à la fois dans leur version originale et dans leur version réduite par un facteur 3.

Ensembles de test

Les images constituant les ensembles de test proviennent de la même campagne d'acquisition que les images utilisées en entraînement. Elles ont cependant été acquises sur une parcelle bien distincte qui a uniquement servi pour la constitution des ensembles de test par perche et par drone. Ces deux ensembles ont été acquis le même jour. Les symptômes de FD sont donc similaires et les conditions d'acquisition très proches – seule l'heure variant. Cette parcelle de Chardonnay comporte plus d'une trentaine de pieds de vigne contaminés par la FD. Elle présente aussi une contamination importante de mildiou, avec des niveaux d'avancement multiples. Des expressions combinées de mildiou et de FD peuvent également être observées sur plusieurs feuilles.

Ensemble d'images acquises par perche Cet ensemble d'images est utilisé pour évaluer les performances des modèles face à une dégradation de résolution spatiale. Il contient 153 images, dont 66 présentent des symptômes de FD. La majeure partie des symptômes de FD est exprimée sur une grande partie du pied de vigne (Figure 5.3 A, B, C) mais certaines images présentent des symptômes sur quelques feuilles uniquement (Figure 5.3 D). Cet ensemble d'images est présenté successivement au modèle sous trois dimensions différentes : 1) *Originale* : 3 976 * 2 652 pixels ; 2) *Intermédiaire* : 1 729 * 1 153 pixels ; 3) *Équivalente à une résolution d'image acquise par drone* : 857 * 571 pixels. L'équivalence a été déterminée en mesurant la taille en pixels d'une trentaine de feuilles sur les images acquises par perche et par drone.



FIGURE 5.3 – Exemples d'images contenues dans l'ensemble de test perche. La première ligne correspond à l'image brute, la seconde à la segmentation manuelle (en clair : les zones présentant de la flavescence dorée).

Ensemble d'images acquises par drone Cet ensemble d'images sert à évaluer les performances des modèles face à des images venant de plateformes différentes de celles utilisées pour acquérir les images d'entraînement – et donc présentant une résolution spatiale et un angle d'acquisition différents. Il contient 27 images acquises par drone (Figure 5.4). Afin de faciliter l'analyse de

ces images, elles ont été découpées par 4, en tuiles de 3976×2652 pixels. Nous obtenons ainsi 108 tuiles, dont 55 contiennent de la FD. Les images ont été acquises par un drone Md4-1000 associé à un appareil photo Sony DSC-RX1RM2 de 42 mégapixels. Le vol a été réalisé en mode manuel, avec une altitude d'environ 10 mètres et une résolution spatiale associée d'environ un millimètre. Cet ensemble d'images est présenté aux modèles dans ses caractéristiques originales, mais aussi dans une version redimensionnée à 1616×1078 pixels, simulant un vol à 25 mètres d'altitude.



FIGURE 5.4 – Exemples d'images contenues dans l'ensemble de test drone. La première ligne correspond à l'image brute, la seconde à la segmentation manuelle (en clair : les zones présentant de la flavescence dorée).

5.2.3 Méthodes

Entraînement

L'architecture de RNC sélectionnée est ResNet-18 (He *et al.*, 2016). Pour les trois variantes de l'ensemble d'entraînement, des apprentissages par affinement sont réalisés. Les poids initiaux sont ceux issus d'un pré-entraînement sur ImageNet et rendus disponibles par PyTorch. Leurs valeurs sont ensuite modifiées par un apprentissage sur l'ensemble d'entraînement pour 15 époques. Tous les RNCs entraînés le sont selon les mêmes hyperparamètres (Tableau 5.1), ceux définis dans le Chapitre 4. Les opérations d'augmentation sont également les mêmes : rotation (probabilité : 0.6), distortions (probabilité : 0.6), retournement (probabilité : 0.8) et modification de la luminosité (probabilité : 0.6). Ces transformations ont été implémentées via la librairie Augmentor³, et elles sont appliquées de façon aléatoire à chaque fois qu'une image est envoyée au réseau. Pour les ensembles d'entraînement des approches 1 et 2, cinq entraînements avec des initialisations différentes sont réalisés. Pour l'ensemble d'origine, le meilleur modèle issu du Chapitre 4 a été directement sélectionné, les modalités d'entraînement étant similaires.

3. <https://augmentor.readthedocs.io/en/master/>

Inférence

Conversion en Réseaux Pleinement Convolutifs Une fois les RNCs entraînés, ceux-ci sont transformés en Réseaux Pleinement Convolutifs (RPC), selon la méthode détaillée par Shelhamer *et al.* (2017). Cette transformation, qui consiste à remplacer la couche pleinement connectée en couche de convolution 1x1, permet de réaliser une segmentation sémantique en inférence. Néanmoins, du fait de l'utilisation de couches de *pooling* (regroupement) et du fait des valeurs de pas et de marge des couches convolutives, la taille de la segmentation obtenue en sortie est bien plus petite que celle de l'image fournie en entrée du RPC. Dans un ResNet, un sous-échantillonnage est effectué à cinq différents niveaux du réseau par une couche convolutive au pas de 2, ce qui conduit à l'obtention d'une sortie 32 fois plus petite que l'image en entrée. Afin de pouvoir superposer l'image en entrée et la prédiction, cette dernière est redimensionnée via une interpolation bilinéaire.

D'un ensemble de test à l'autre, seul un hyperparamètre a été modifié : la taille des régions de *pooling* dans la dernière couche d'*average pooling*. Cette couche opère un sous-échantillonnage en divisant la carte de caractéristiques en entrée en sous-zones rectangulaires. La valeur moyenne est ensuite calculée et reportée sur la carte de caractéristiques de sortie (Goodfellow *et al.*, 2016). Dans le cadre de l'inférence en RPC, cette opération permet de lisser la prédiction, impactant donc la finesse de la segmentation. Plus la taille des régions de *pooling* est élevée, plus la résolution de la prédiction est grossière. Afin d'avoir une concordance entre la taille des zones de *pooling* et la taille de l'image présentée en inférence, les tailles de zones *pooling* ont été fixées comme suit : 1.1) *Perche original* : 13×13 pixels ; 1.2) *Perche intermédiaire* : $9 * 9$ pixels ; 1.3) *Perche équivalent drone* : $5 * 5$ pixels ; 2.1) *Drone original* : $11 * 11$ pixels ; 2.2) *Drone équivalent 25 mètres* : $5 * 5$ pixels. Même si les ensembles perche original et drone original ont des images de même dimension, une taille de région de *pooling* plus élevée a été choisie pour l'ensemble perche afin de limiter le nombre de faux positifs, plus susceptibles d'apparaître dans ces images très détaillées.

Métriques Plusieurs métriques sont utilisées pour comparer les différents modèles entre eux. Tout d'abord, afin de connaître le nombre de prédictions correctes et incorrectes pour chacune des classes, le nombre de faux positifs et de faux négatifs ont été relevés. Un exemple positif (de la classe

TABLEAU 5.1 – Hyperparamètres utilisés pour entraîner les modèles basés sur l'architecture ResNet-18.

Hyperparamètre	Équivalent anglais	Valeur
Taux d'apprentissage	<i>Learning rate</i>	1e-04
Planificateur	<i>Learning rate scheduler</i>	Pas : 5, Gamma : 0.7
Optimiseur	<i>Optimizer</i>	Adam
Fonction de perte	<i>Loss function</i>	Entropie croisée
Taille du batch	<i>Batch size</i>	64
Nombre d'époques	<i>Epochs</i>	15

"FD") est défini comme vrai positif quand il y a une intersection non nulle entre les segments prédits et l'annotation. Cette absence de seuil minimum d'intersection permet de maximiser la sensibilité du modèle à la détection de symptômes de FD. Un exemple négatif (de la classe "Non FD") est défini comme vrai négatif quand l'intégralité de l'image a correctement été prédite comme négative.

Ensuite, afin d'obtenir une vision plus complète des performances de ces modèles, la précision (Eq. 5.1), le rappel (Eq. 5.2) et le F1-score (Eq. 5.3) sont calculés. C'est ce dernier indice qui est retenu pour classer les modèles entre eux.

$$Précision = \frac{VP}{VP + FP} \quad (5.1)$$

où VP (Vrai Positif) correspond au nombre total d'images positives (de la classe "FD") bien identifiées en tant que telles, et où FP (Faux Positif) correspond au nombre total d'images contenant au moins un faux positif.

$$Rappel = \frac{VP}{VP + FN} \quad (5.2)$$

où VP correspond au nombre total d'images positives (de la classe "FD") bien identifiées en tant que telles, et où FN (Faux négatif) correspond au nombre total d'images positives (de la classe "FD") non prédites comme telles.

$$F1 = 2 * \frac{Précision * Rappel}{Précision + Rappel} \quad (5.3)$$

Enfin, afin d'évaluer la qualité de la segmentation, l'intersection sur union moyenne (*mean Intersection over Union* – mIoU) pour chaque classe est calculée (Eq. 5.4).

$$mIoU_{AB} = \frac{A \cap B}{A \cup B} \quad (5.4)$$

où A correspond au nombre de pixels prédits pour une classe cible et B correspond au nombre de pixels réels de la classe cible.

5.3 Résultats

5.3.1 Phase d'entraînement

Une convergence rapide des modèles peut être observée pour les deux approches et ce, pour les cinq différentes initialisations de l'entraînement (Figure 5.5). A la fin de la première époque, les taux de bonnes prédictions sur l'ensemble des patches de validation sont compris entre 97,14 % et 98,9 % pour la première approche, et entre 98,46 % et 99,17 % pour la seconde. Le taux le plus élevé est obtenu à la 13ème époque dans les deux approches, avec 99,62 % pour la première, et 99,96 % pour la seconde. L'époque permettant d'obtenir le taux de bonnes prédictions le plus élevé pour

chacune des cinq initialisations est retenue puis convertie en RPC pour l'inférence sur les ensembles d'évaluation. Pour chaque approche, cinq modèles sont ainsi évalués.

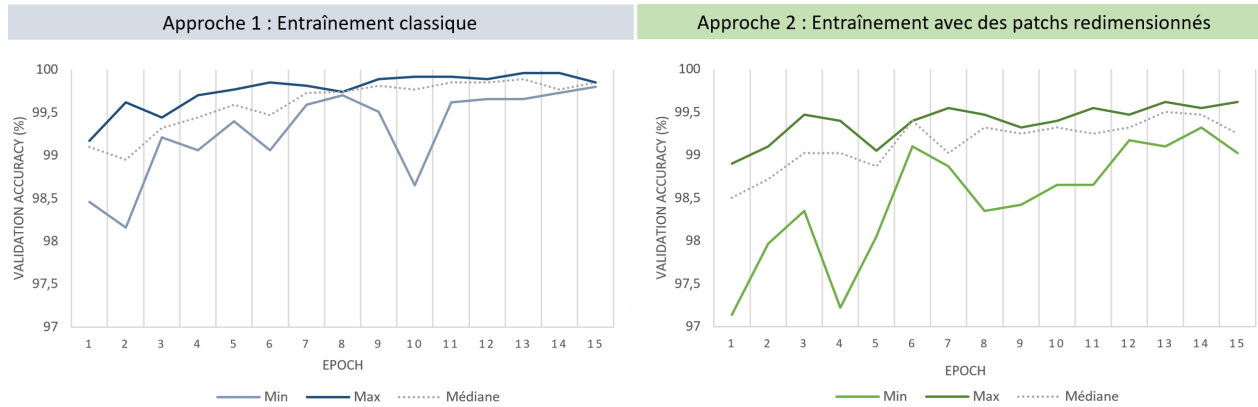


FIGURE 5.5 – Précisions minimales, maximales et médianes obtenues sur l'ensemble de validation pour les approches 1 et 2, et leurs 5 initialisations.

5.3.2 Performances des modèles face à une dégradation de résolution spatiale : inférence sur les images acquises par perche

Approche 1 : Modèles issus d'un entraînement sur un jeu de données enrichi

Sur les images aux dimensions originales, le F1-score s'étend de 0,79 à 0,91 (Tableau 5.2). Pour l'ensemble des initialisations, le rappel est très élevé – de 0,98 à 1. La précision est quant à elle beaucoup plus contrastée, allant de 0,66 à 0,84.

Sur les images de dimension intermédiaire, peu de variabilité est à observer sur le F1-score, qui est compris entre 0,95 et 0,97. Contrairement aux inférences sur la dimension originale, le nombre de faux positifs est limité et surtout proche d'une initialisation à l'autre, avec une précision allant de 0,94 à 0,96. Le modèle au F1-score le plus élevé atteint une valeur de rappel de 1.

Sur les images aux dimensions équivalentes à celles d'images acquises par drone, le F1-score s'étend de 0,74 à 0,97. La différence entre ces modèles issus des différentes initialisations se trouve avant tout dans le nombre d'images contenant des faux positifs.

Toutes dimensions confondues, le F1-score le plus élevé est obtenu par l'inférence sur les images aux dimensions intermédiaires, avec 0,97. Ce résultat souligne notamment qu'une quantité plus importante de détails dans une image n'est ici pas liée à de meilleures prédictions. Elle conduit même à un nombre conséquent de faux positifs. Néanmoins, ces faux positifs, bien que nombreux, ne concernent que peu de pixels, comme le reflète le mIoU qui demeure élevé pour la classe "Non-FD". Ces fausses détections apparaissent notamment sur des feuilles aux nervures courbées et au bord peu dentelé (Figure 5.6, A) – deux caractéristiques qui avaient été associées à la classe "FD" dans le Chapitre 4 – mais aussi au niveau de feuilles floues de premier plan (Figure 5.6, B) et de rameaux

TABLEAU 5.2 – Synthèse des résultats d'inférence obtenus avec les modèles entraînés avec l'approche 1 sur les trois dimensions d'image étudiées. Des cinq initialisations d'entraînement réalisées, seuls les modèles obtenant la valeur minimale, médiane et maximale de F1-Score sont présentés. Les nombres en gras correspondent aux valeurs de précision, rappel ou F1-score les plus élevées. VP = Vrai Positif, P = Positif, FP = Faux Positif, N = Négatif, mIoU = *mean Intersection over Union*.

	Dimension originale			Dimension intermédiaire			Dimension équiv. drone		
	<i>Min</i>	<i>Médiane</i>	<i>Max</i>	<i>Min</i>	<i>Médiane</i>	<i>Max</i>	<i>Min</i>	<i>Médiane</i>	<i>Max</i>
ID modèle	5	2	3	5	4	3	3	2	5
VP (pour 66 P)	65	66	65	63	64	66	64	62	62
FP (pour 87 N)	33	24	12	4	3	4	23	7	2
Précision	0,66	0,73	0,84	0,94	0,96	0,94	0,74	0,90	0,97
Rappel	0,98	1,00	0,98	0,95	0,97	1,00	0,97	0,94	0,94
F1-Score	0,79	0,85	0,91	0,95	0,96	0,97	0,84	0,92	0,95
mIoU									
<i>FD</i>	0,54	0,56	0,55	0,51	0,51	0,48	0,46	0,49	0,47
<i>Non-FD</i>	0,89	0,89	0,90	0,86	0,86	0,84	0,84	0,87	0,88

(Figure 5.6, C et D). Enfin, une spécialisation des modèles pour une résolution donnée peut être constatée. En effet, si l'on s'intéresse aux ID des modèles retenus pour les F1-score minimum, médian et maximum, on observe que le modèle obtenant le score le plus élevé pour les dimensions originale et intermédiaire est l'initialisation n° 3. Pour la dimension équivalente drone, cette initialisation obtient les moins bons résultats, avec un F1-score de 0,84, contre 0,95 pour l'initialisation n° 5, qui obtenait quant à elle les scores les plus faibles sur les deux autres dimensions. Le modèle au moins bon score pour les deux premières dimensions devient ainsi le meilleur pour la troisième, et vice versa.

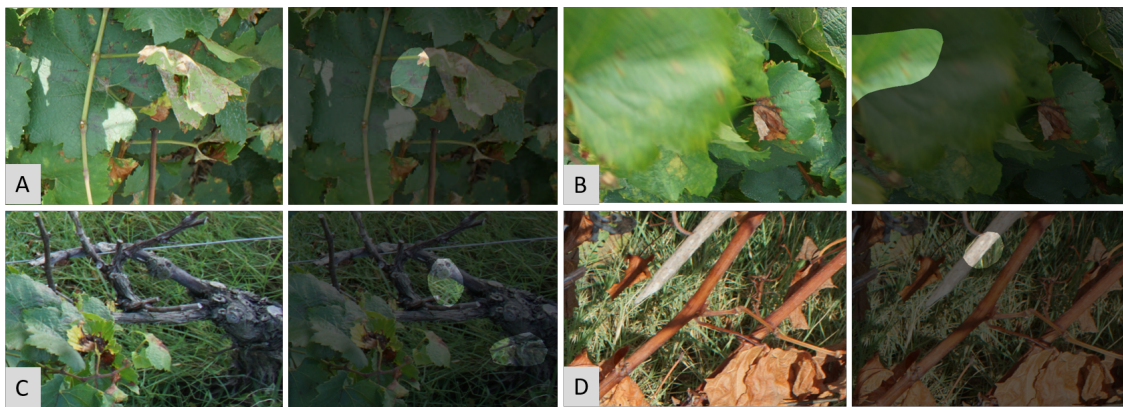


FIGURE 5.6 – Exemples de fausses détections de flavescence dorée obtenues lors de l'inférence du meilleur modèle de l'approche 1 sur les images de dimension originale. Les zones sombres sont celles prédites comme "Non flavescence", les zones claires celles prédites comme "Flavescence".

Approche 2 : Modèles issus d'un entraînement avec des patches redimensionnés

Sur les images aux dimensions originales, le F1-score s'étend de 0,68 à 0,93 (Tableau 5.3). Pour l'ensemble des initialisations, le rappel est très élevé – de 0,98 à 1. Comme pour la première approche, le nombre de faux positifs sur cette dimension est important, avec un score de précision allant de 0,52 à 0,90. La valeur de mIoU reste élevée pour la classe "Non-FD", entre 0,88 et 0,91, soulignant de nouveau la petite dimension de ces faux positifs.

Sur les images de dimension intermédiaire, les valeurs de F1-score sont très proches d'une initialisation à l'autre, avec des valeurs allant de 0,94 à 0,98. Le nombre d'images contenant des faux positifs est quant à lui beaucoup plus restreint que pour la dimension originale. La précision s'étend de 0,93 à 0,98.

Sur les images aux dimensions équivalentes à celles d'images acquises par drone, le F1-score s'étend de 0,89 à 0,94. Les valeurs de rappel, qui vont de 0,91 à 0,95, constituent les plus faibles obtenues des trois dimensions – les inférences sur les dimensions originale et intermédiaire ayant des valeurs de rappel allant de 0,95 à 1.

Comme pour la première approche, toutes dimensions confondues, le F1-score le plus élevé est obtenu par l'inférence sur les images aux dimensions intermédiaires, avec 0,98. On observe également la même dynamique de performance entre les différentes initialisations de modèles. L'initialisation n° 5, qui obtient les F1-score les plus faibles pour les inférences sur les images de dimensions originale et intermédiaire, est celui qui obtient le score le plus élevé lors de l'inférence sur les images aux dimensions équivalentes drone. La dynamique de classement inverse est observable pour l'initialisation n° 1. Quant à la qualité de la segmentation pour la classe "FD", elle diminue progressivement avec la diminution de la taille de l'image d'inférence, avec un écart maximal de 0,09, soit du même ordre de grandeur que celui relevé dans la première approche.

TABLEAU 5.3 – Synthèse des résultats d'inférence obtenus avec la seconde approche sur les trois dimensions d'image étudiées. Des cinq initialisations d'entraînement réalisées, seuls les modèles obtenant la valeur minimale, médiane et maximale de F1-Score sont présentés. Les nombres en gras correspondent aux valeurs de précision, rappel ou F1-score les plus élevées. VP = Vrai Positif, P = Positif, FP = Faux Positif, N = Négatif, mIoU = *mean Intersection over Union*.

	Dimension originale			Dimension intermédiaire			Dimension équiv. drone		
	Min	Médiane	Max	Min	Médiane	Max	Min	Médiane	Max
ID modèle	5	2	1	5	3	1	1	3	5
VP (pour 66 P)	65	66	63	63	63	64	60	62	63
FP (pour 87 N)	61	21	7	5	2	1	9	7	5
Précision	0,52	0,76	0,90	0,93	0,97	0,98	0,87	0,90	0,93
Rappel	0,98	1,00	0,95	0,95	0,95	0,97	0,91	0,94	0,95
F1-Score	0,68	0,86	0,93	0,94	0,96	0,98	0,89	0,92	0,94
mIoU									
<i>FD</i>	0,55	0,56	0,53	0,52	0,5	0,51	0,48	0,49	0,47
<i>Non-FD</i>	0,88	0,90	0,91	0,86	0,85	0,86	0,87	0,86	0,86

Comparaison des approches

La comparaison des résultats d'inférence des modèles de l'approche 1, de l'approche 2 et du Chapitre 4 montre qu'aucun modèle n'est le meilleur pour l'ensemble des trois dimensions (Tableau 5.4).

Pour la dimension originale, le F1-score le plus élevé est obtenu par le modèle du Chapitre 4, avec 0,98. Il se démarque par sa capacité à avoir peu de faux positifs (1,15 %), contrairement au modèle de l'approche 1 (8,05 %) et surtout à celui de l'approche 2 (13,79 %) (Figure 5.7). Ce jeu d'entraînement a donc conduit à des modèles tout aussi sensibles mais moins spécifiques sur des images de cette dimension.

TABLEAU 5.4 – Comparaison des résultats d'inférence obtenus avec le modèle au F1-score le plus élevé pour les deux approches ainsi que le modèle développé dans le Chapitre 4. Les nombres en gras correspondent aux valeurs de précision, rappel ou F1-score les plus élevées. VP = Vrai Positif, P = Positif, FP = Faux Positif, N = Négatif, mIoU = *mean Intersection over Union*.

	Dimension originale			Dimension intermédiaire			Dimension équiv. drone		
	App. 1	App. 2	Ch. 4	App. 1	App. 2	Ch. 4	App. 1	App. 2	Ch. 4
ID modèle	3	1	N/A	3	1	N/A	5	5	N/A
VP (pour 66 P)	65	63	65	66	64	64	62	63	62
FP (pour 87 N)	12	7	1	4	1	8	2	5	29
Précision	0,84	0,90	0,98	0,94	0,98	0,89	0,97	0,93	0,68
Rappel	0,98	0,95	0,98	1,00	0,97	0,97	0,94	0,95	0,94
F1-Score	0,91	0,93	0,98	0,97	0,98	0,93	0,95	0,94	0,79
mIoU									
FD	0,55	0,53	0,53	0,48	0,51	0,52	0,47	0,47	0,45
Non-FD	0,90	0,91	0,91	0,84	0,86	0,87	0,88	0,86	0,86

Pour la dimension intermédiaire, une dynamique différente est observée par rapport à la dimension originale. On constate une diminution importante du nombre de faux positifs pour les approches 1 et 2, avec une précision qui passe de 0,84 à 0,94 pour la première et de 0,90 à 0,98 pour la seconde. Au contraire, pour le modèle issu du Chapitre 4, le nombre de faux positifs augmente, avec une précision passant de 0,98 à 0,89. Le nombre de vrais positifs augmente légèrement (+1) pour les approches 1 et 2. Leurs F1-scores s'élèvent respectivement à 0,97 et à 0,98.

Sur les images aux dimensions équivalentes à celles d'images acquises par drone, on constate également cette proximité des résultats des approches 1 et 2, avec un F1-score de 0,95 et 0,94. Par rapport aux inférences sur les autres dimensions, on observe une diminution globale du rappel, qui se maintient toutefois entre 0,94 et 0,95. Pour le modèle du Chapitre 4, le F1-score tombe à 0,79 du fait d'un nombre conséquent de faux positifs (33,33 %).

Dans l'ensemble, les modèles conservent une sensibilité importante au fil des diminutions de résolution des images de test, avec un score de rappel de 0,94 au minimum. Cette sensibilité est très pertinente pour notre thématique, une omission de FD étant plus coûteuse qu'une fausse détection. Elle est également encourageante quant à l'application des modèles sur des images possédant une plus faible résolution, comme les images acquises par drone. Les zones de FD qui n'ont pas été détectées sont

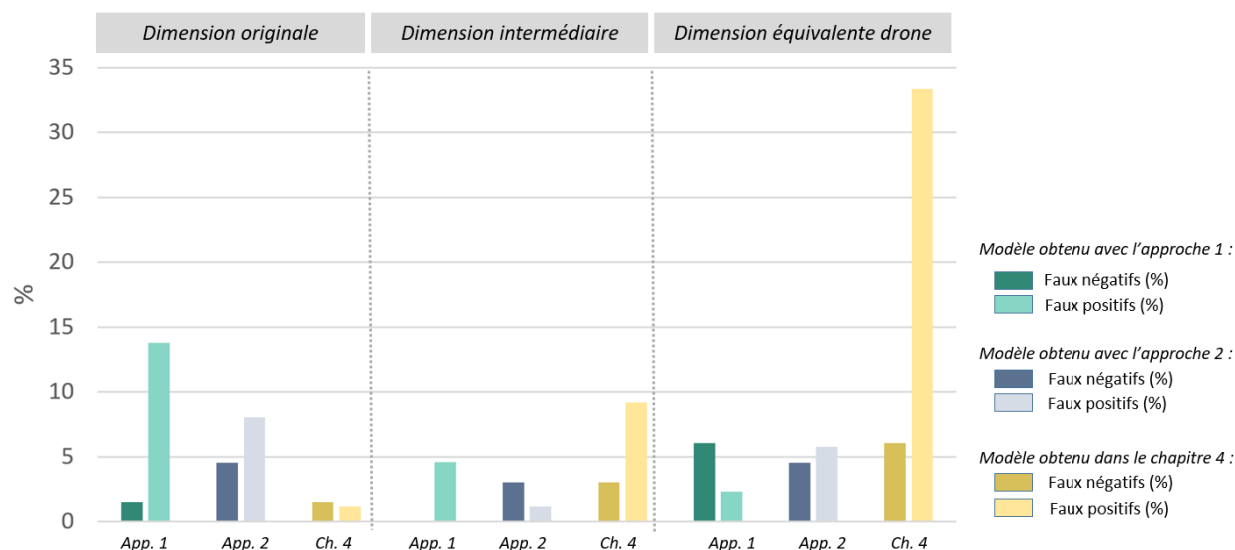


FIGURE 5.7 – Pourcentages de faux négatifs et de faux positifs obtenus lors de l'inférence sur l'ensemble des images acquises par perche avec les modèles issus de l'approche 1, de l'approche 2 et du Chapitre 4, pour les 3 dimensions étudiées.

récurrentes d'un modèle à l'autre. Pour les trois approches, ce sont sensiblement les mêmes images qui sont à l'origine des faux négatifs (Figure 5.8). Deux points communs peuvent être relevés entre ces images : 1) les zones de FD sont de petite taille, exprimées sur quelques feuilles seulement, 2) les feuilles expriment à la fois un symptôme typique de la FD, l'enroulement, mais est également marquée par des taches de mildiou (Figure 5.8, B, C, D). Deux des images contenant des faux négatifs correspondent à la même plante photographiée sous deux angles différents (Figure 5.8, C, D). Le modèle de l'approche 2 est parvenu à reconnaître de la FD sur l'une des deux images uniquement, soulignant la sensibilité au point de vue.

Quant à la spécificité, celle-ci fluctue d'un modèle à l'autre et d'une dimension à l'autre. Alors que le modèle du Chapitre 4 permet de minimiser le nombre de faux positifs pour les images de dimension originale, le nombre de faux positifs est plus faible avec les approches 1 et 2 pour les deux autres dimensions. L'enrichissement du jeu d'entraînement avec des ombres, des effets foliaires complexes, des structures de palissage et des végétaux autres que la vigne accroît donc la spécificité du modèle lors de l'analyse d'une image aux dimensions intermédiaires et équivalentes à celles d'une image drone. Par ailleurs, les résultats des approches 1 et 2 sont très proches. L'ajout de patches sous-dimensionnés ne semble donc pas apporter une meilleure capacité d'analyse au modèle pour les résolutions spatiales les plus faibles.

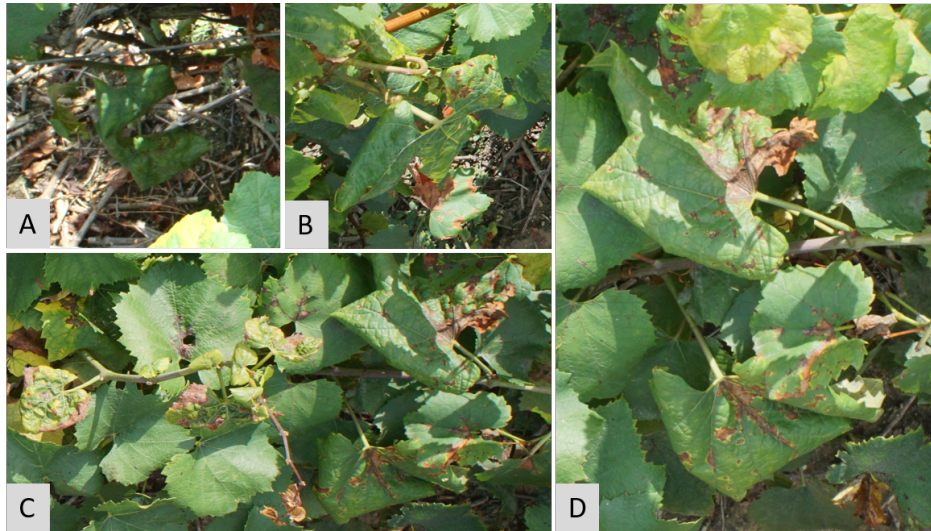


FIGURE 5.8 – Feuilles présentant des symptômes de flavescence dorée n’ayant pas été détectés lors de l’inférence sur une dimension similaire à celle d’images drone avec les trois approches (A, B, C), et avec l’approche 1 et celle du Chapitre 4 (D).

5.3.3 Performances des modèles face des images venant de différentes plateformes : inférence sur les images acquises par drone

Approche 1 : Modèles issus d’un entraînement sur un jeu de données enrichi

Les prédictions obtenues sur les images drone sont de qualité inférieure à celles obtenues sur les images perche aux dimensions équivalentes (Tableau 5.5). Le F1-score s’élève en effet au maximum à 0,61, contre 0,95 sur les images perche. On observe à la fois une diminution du rappel (de 0,78 contre 0,97), mais surtout de la précision (de 0,50 contre 0,94).

Sur les images simulant un vol drone à 25 mètres, les prédictions sont d’autant plus de faible qualité, avec un F1-score de 0,35 au maximum. Le nombre de vrais positifs est faible, avec un rappel s’élevant seulement à 0,31.

Approche 2 : Modèles issus d’un entraînement avec des patchs redimensionnés

Avec le modèle de la seconde approche, les prédictions obtenues sur les images drone sont également de qualité inférieure à celles obtenues sur les images perche aux dimensions équivalentes (Tableau 5.6). Le F1-score s’élève en effet au maximum à 0,76, contre 0,94 sur les images perche. On observe une différence assez importante entre les performances des différentes initialisations, à la fois en terme de rappel et de précision.

Sur les images simulant un vol drone à 25 mètres, le F1-score s’élève à 0,31 au maximum. Là encore, on observe une importante variabilité entre les initialisations.

TABLEAU 5.5 – Synthèse des résultats d’inférence obtenus avec la première approche sur les deux dimensions d’images acquises par drone étudiées. Des cinq initialisations d’entraînement réalisées, seuls les modèles obtenant la valeur minimale, médiane et maximale de F1-Score sont présentés. Les nombres en gras correspondent aux valeurs de précision, rappel ou F1-score les plus élevées. VP = Vrai Positif, P = Positif, FP = Faux Positif, N = Négatif, mIoU = *mean Intersection over Union*.

	Dimension originale			Dimension équivalente 25m		
	<i>Min</i>	<i>Médiane</i>	<i>Max</i>	<i>Min</i>	<i>Médiane</i>	<i>Max</i>
ID modèle	4	2	1	5	1	3
VP (pour 55 P)	39	39	43	5	9	17
FP (pour 53 N)	49	40	42	2	5	25
Précision	0,44	0,49	0,50	0,71	0,64	0,40
Rappel	0,71	0,71	0,78	0,10	0,16	0,31
F1-Score	0,54	0,58	0,61	0,16	0,26	0,35
mIoU						
<i>FD</i>	0,23	0,24	0,23	0,16	0,16	0,18
<i>Non-FD</i>	0,96	0,97	0,97	0,99	0,99	0,98

TABLEAU 5.6 – Synthèse des résultats d’inférence obtenus avec la seconde approche sur les deux dimensions d’image acquises par drone étudiées. Des cinq initialisations d’entraînement réalisées, seuls les modèles obtenant la valeur minimale, médiane et maximale de F1-Score sont présentés. Les nombres en gras correspondent aux valeurs de précision, rappel ou F1-score les plus élevées. VP = Vrai Positif, P = Positif, FP = Faux Positif, N = Négatif, mIoU = *mean Intersection over Union*

	Dimension originale			Dimension équivalente 25m		
	<i>Min</i>	<i>Médiane</i>	<i>Max</i>	<i>Min</i>	<i>Médiane</i>	<i>Max</i>
ID modèle	2	3	5	1	2	4
VP (pour 55 P)	28	39	38	5	9	17
FP (pour 53 N)	11	24	7	17	19	37
Précision	0,71	0,62	0,84	0,23	0,32	0,31
Rappel	0,51	0,71	0,70	0,10	0,16	0,31
F1-Score	0,59	0,66	0,76	0,13	0,22	0,31
mIoU						
<i>FD</i>	0,27	0,25	0,27	0,15	0,17	0,16
<i>Non-FD</i>	0,98	0,98	0,99	0,98	0,98	0,97

Comparaison des approches

Le modèle permettant d’obtenir les meilleures prédictions sur les images drone originales est celui de l’approche 2 (Tableau 5.7). Les résultats obtenus sur ces images sont globalement plus faibles que ceux obtenus pour l’analyse des images perche à une résolution équivalente, avec un écart de 0,19 entre les F1-score des meilleurs modèles. Au-delà de la résolution, d’autres caractéristiques d’acquisition sont impactées avec le passage à une plateforme drone, principalement l’angle d’acquisition. Même si, sur nos images drone, les feuilles de vigne ont tendance à se positionner horizontalement au-dessus de la canopée, on y observe plus d’effets d’ombre et d’enchevêtrement de feuilles, des sources de faux positifs. Les scènes observées par perche et par drone comportent donc de nombreuses différences, pouvant influencer de manière trop conséquentes les statistiques de l’image pour permettre l’analyse des images drone par un modèle entraîné sur des images perche.

TABLEAU 5.7 – Comparaison des résultats d’inférence obtenus avec le modèle au F1-score le plus élevé pour les deux approches ainsi que le modèle développé dans le Chapitre 4 sur les images acquises par drone. Les nombres en gras correspondent aux valeurs de précision, rappel ou F1-score les plus élevées. VP = Vrai Positif, P = Positif, FP = Faux Positif, N = Négatif, mIoU = *mean Intersection over Union*.

	Dimension originale			Dimension équivalente 25m		
	App. 1	App. 2	Ch.4	App. 1	App. 2	Ch. 4
ID modèle	1	5	N/A	3	4	N/A
VP (pour 55 P)	43	38	42	17	17	4
FP (pour 53 N)	42	7	53	25	37	12
Précision	0,50	0,84	0,44	0,40	0,31	0,25
Rappel	0,78	0,70	0,76	0,31	0,31	0,10
F1-Score	0,61	0,76	0,56	0,35	0,31	0,11
mIoU						
<i>FD</i>	0,23	0,27	0,20	0,18	0,16	0,1
<i>Non-FD</i>	0,97	0,99	0,95	0,98	0,97	0,98

L’enrichissement du jeu d’entraînement a justement conduit à un taux de faux positifs plus faible (100 % pour le modèle du Chapitre 4 contre 79 % pour celui de l’approche 1), mais c’est surtout la dégradation de dimension de patchs lors de l’entraînement qui permet d’obtenir le taux de faux positifs le plus faible (13 %) (Figure 5.9). Cependant, l’approche 2 conduit au modèle au nombre de faux négatifs le plus élevé (31 %). Les zones de FD qui n’ont pas été repérées sont de nature multiple : des zones à la contamination étendue ou au contraire localisée, des zones exprimant un fort enroulement, d’autres exprimant plutôt un jaunissement foliaire ou encore exprimant à la fois du mildiou et de la FD (Figure 5.10). Il est donc difficile de dresser un profil type de ces faux négatifs. Ainsi, aucun des modèles générés n’est adapté pour repérer la FD sur les images acquises par drone. L’application des modèles sur des images simulant un vol drone à 25 mètres conduit à des résultats de très faible qualité, avec un F1-score maximal de 0,35, obtenu avec la première approche. Pour les trois modèles, le taux de faux négatifs est très élevé (de 69 % à 92 %). Même dans les images où de

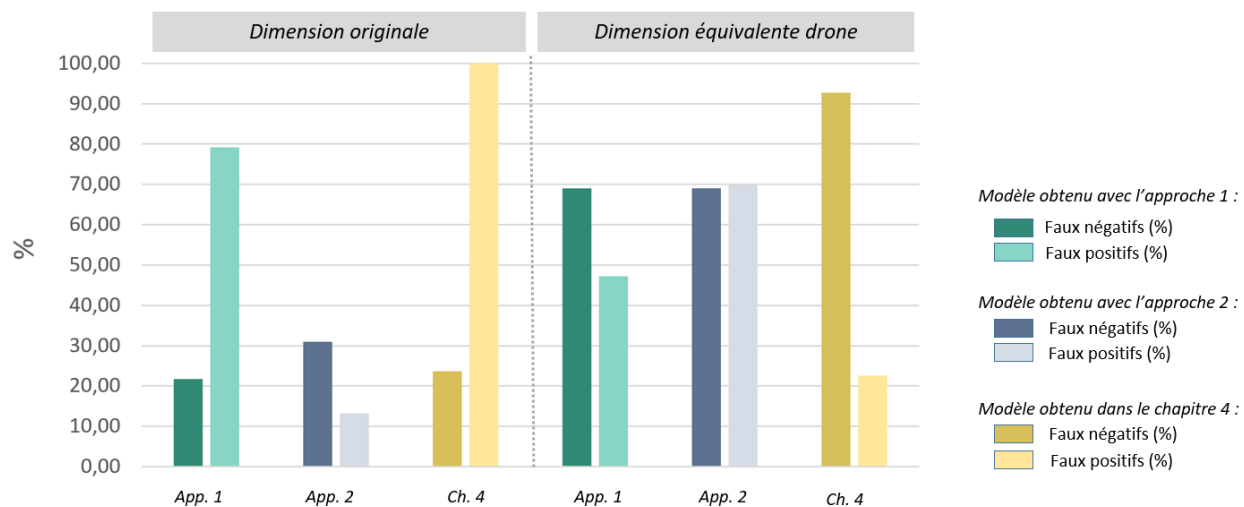


FIGURE 5.9 – Pourcentages de faux négatifs et de faux positifs lors de l'inférence sur les images acquises par drone avec les modèles issus de l'approche 1, de l'approche 2 et du Chapitre 4.



FIGURE 5.10 – Exemples de zones de flavescence dorée non repérées par le modèle de l'approche 2 sur les images acquises par drone.

bonnes identifications occurrent, ce n'est pas l'intégralité des zones de FD qui sont repérées (Figure 5.11), d'où une valeur d'IoU faible pour la classe "FD".

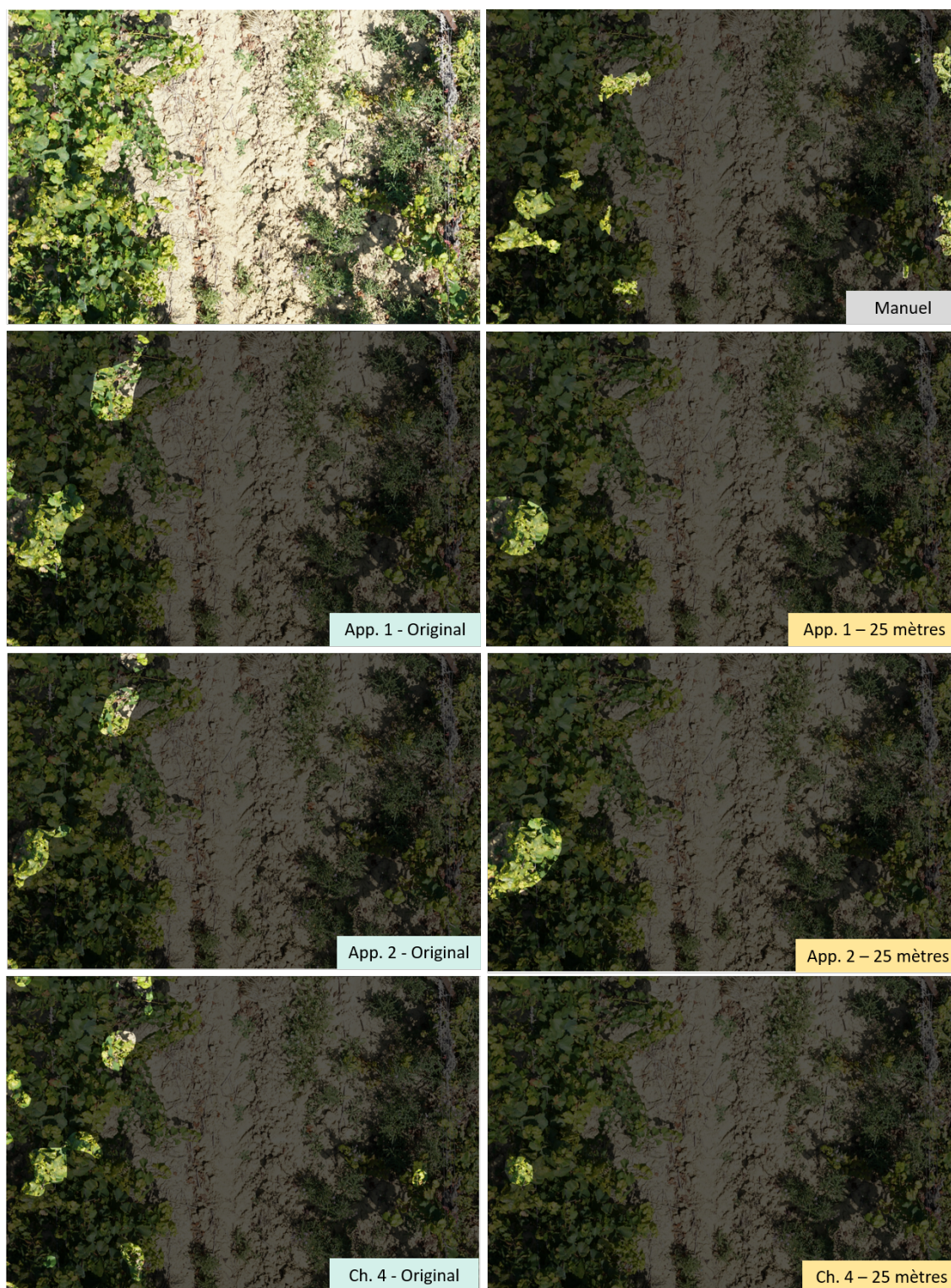


FIGURE 5.11 – Inférence des modèles issus des deux approches étudiées et du cas d'étude développé dans le Chapitre 4 sur une image drone, en dimensions originale et équivalente à un vol à 25 mètres d'altitude.

5.4 Discussion et conclusion

5.4.1 Les RNCs face au changement de résolution spatiale

Impact de la résolution spatiale

Dans cette étude, nous avons pu observer qu'en inférence, la résolution spatiale la plus élevée n'apportait pas les meilleurs résultats de segmentation. Sur les ensembles d'images acquises par perche, les meilleures prédictions ont en effet été obtenues avec la dimension intermédiaire et non pas avec la dimension originale. Cette performance pourrait être expliquée par le fait que l'entraînement est réalisé sur des patches aux dimensions réduites à $224 * 224$ pixels, résultant en un modèle plus adapté à l'analyse d'images à plus faible résolution. Cet aspect est avantageux pour l'inférence, qui est alors moins coûteuse en ressources de calcul et plus rapide. La dimension des images présentées en inférence peut donc fortement impacter les performances. Elle doit donc être considérée comme un paramètre à adapter, notamment en fonction de la thématique étudiée et de la taille des images d'entraînement, dans l'optique de faire concorder en entraînement et en inférence l'échelle observée de l'objet d'intérêt. Pour améliorer l'analyse d'images à plus faible résolution (comme celles équivalentes à des résolutions drone dans notre cas), un sur-échantillonnage pourrait d'ailleurs être envisagé afin de faire concorder artificiellement l'échelle d'observation entre l'entraînement et l'inférence. Une autre technique, fréquemment utilisée pour améliorer les performances est de présenter en inférence successivement la même image sous différentes résolutions puis de moyenner les prédictions (Simonyan et Zisserman, 2015; He *et al.*, 2016; Touvron *et al.*, 2019).

La résolution spatiale est également un paramètre à considérer lors de l'entraînement de classifieur. Sabottke et Spieler (2020) ont étudié l'effet de la résolution des images utilisées pour entraîner des RNCs pour l'analyse de radiographies. Les meilleures performances ont été obtenues avec des images aux dimensions comprises entre $256 * 256$ pixels et $448 * 448$ pixels. Ils ont également constaté que la taille idéale variait selon la maladie à diagnostiquer. Certaines des pathologies étudiées ont plus fortement bénéficié de l'augmentation de la résolution pour leur détection, notamment les nodules pulmonaires – une lésion de petite taille, environ 3cm – ou encore, les emphysèmes qui s'expriment de façon diffuse sur la radiographie. Kawahara et Hamarneh (2016) ont, quant à eux, fait le choix de l'entraînement multi-résolution pour classifier des lésions dermatologiques. Simultanément, deux versions de différente résolution de la même image sont envoyées au travers de deux branches d'analyse ne partageant pas les mêmes paramètres. Les réponses de ces deux branches sont ensuite combinées pour générer la prédiction finale. Une autre manière de réaliser un entraînement multi-résolution est de s'émanciper de l'obligation de la taille d'entrée fixe, qui est requise du fait la présence de la couche pleinement connectée en fin de réseau. He et al. (2015) ont proposé l'ajout d'une couche de *pooling* générant une sortie de la taille requise pour alimenter le classifieur – *the spatial pyramid pooling*. La librairie PyTorch propose une couche de *pooling* à la

fonction similaire : `AdaptiveAvgPool2d`⁴. L'avantage de telles opérations est de réaliser l'agrégation des caractéristiques à un niveau profond du réseau, et non pas dès l'entrée, permettant ainsi de fournir au modèle des images plus détaillées et donc potentiellement plus propices à l'analyse.

Ainsi, autant en entraînement qu'en inférence, la résolution spatiale est un paramètre qu'il est assurément intéressant d'explorer pour tirer les meilleures performances d'un modèle.

Impact du jeu de données

Dans cette étude, nous avons également pu observer l'impact du jeu de données utilisé en entraînement. Les deux variantes avec plus d'annotations n'ont pas permis une amélioration des résultats sur les images de dimension originale. Ainsi, avec un ensemble d'entraînement élargi et avec des exemples choisis en réponse à des biais observés, les résultats ne se sont pas améliorés, et même au contraire, dégradés. Cela souligne la difficulté de concevoir l'ensemble d'entraînement le plus adapté à l'objet d'étude. Trouver au travers de l'acquisition et de l'annotation de données les caractéristiques les plus représentatives d'une situation donnée est loin d'être une tâche triviale mais elle est au coeur de la capacité de généralisation du modèle (Corfield, 2009). Toutefois, dans notre cas d'étude, ces mêmes annotations ont eu un impact positif sur la capacité d'analyse d'images à plus faible résolution. L'ajout d'annotations reflétant des structures plus typiquement observées sur des images drone était donc une approche pertinente. Elle a également permis d'obtenir les meilleurs résultats lors de l'analyse des images acquises par drone, même si ces derniers sont largement insuffisants pour envisager une quelconque application réelle à ce stade-ci. Ainsi, avec ces nouvelles annotations, nous avons abouti à modèle plus robuste à l'échelle, mais avec en contre partie une perte de performance sur les images en dimension originale. En pratique, cette spécialisation sur une échelle n'est pas problématique, le besoin se trouvant plutôt dans un modèle adapté à chaque configuration. Il est également à noter que l'architecture et les valeurs d'hyperparamètres ont été choisis en fonction des caractéristiques de l'ensemble de données du Chapitre 4. Des choix plus adaptés aux deux autres ensembles conduiraient à d'autres performances, possiblement meilleures.

5.4.2 Les RNCs, d'une plateforme d'acquisition à l'autre

Dans cette étude, le modèle entraîné sur des images acquises par perche n'est pas parvenu à analyser convenablement les images acquises par drone, le changement de plateforme semblant avoir conduit à une divergence de données trop importante. Cette question récurrente de la capacité restreinte des modèles d'AP à généraliser sur des images aux caractéristiques différentes de celles utilisées en entraînement est au centre du champ de recherche de l'adaptation de domaine. L'adaptation de domaine est un cas particulier d'apprentissage par transfert, qui utilise des données annotées dans un domaine source pour une application dans un domaine cible, pour lequel peu ou pas de données sont disponibles. Dans le cas de l'adaptation de domaine, les domaines source et cible

4. <https://pytorch.org/docs/stable/generated/torch.nn.AdaptiveAvgPool2d.html>

sont connexes mais leur distribution est différente (Sun, 2013; Wang et Deng, 2018). Même pour le cas spécifiques de l’AP, plusieurs approches existent pour résoudre cette problématique. La synthèse réalisée par Wang et Deng (2018) souligne la diversité des approches possibles pour l’AP, allant de l’affinement de l’apprentissage à l’utilisation de réseaux adverses génératifs (*generative adversarial network*). Récemment, Sun *et al.* (2019) ont également proposé une méthode d’adaptation de domaine par apprentissage auto-supervisé (*self-supervised learning*). Dans cette approche, le modèle apprend ses représentations en résolvant des tâches prétextes, telles que de la colorisation, des puzzles ou la reconstitution d’une partie d’image. Le point commun entre ces tâches est que leurs annotations peuvent être générées automatiquement. Il est donc possible d’exploiter de grands ensembles d’images non annotés manuellement. Les représentations apprises sont ensuite utilisées pour résoudre le problème d’intérêt. Certains travaux ont observé qu’un pré-entraînement par apprentissage auto-supervisé permettait d’obtenir de meilleurs résultats qu’un pré-entraînement classique utilisant ImageNet et ses annotations en apprentissage supervisé (Jing et Tian, 2019), mais aussi une plus grande robustesse face à des exemples adversariaux et à des défauts présents dans les images (e.g. bruits, flou) (Hendrycks *et al.*, 2019).

L’apprentissage actif (*active learning*) permet lui aussi de tirer profit de larges ensembles de données, en ayant cette fois recours à peu d’annotations. Cette approche se base sur l’hypothèse que le modèle peut apprendre d’aussi bonnes, voire meilleures, représentations si on lui laisse la sélection des images à annoter. Dans ces systèmes, le modèle envoie à un annotateur humain les images pour lesquelles il obtient des prédictions au taux de confiance faible. Ainsi, seuls ces cas difficiles doivent être annotés, évitant la redondance de l’annotation de cas déjà bien gérés par le modèle (Settles, 2009). En utilisant de l’apprentissage actif pour la segmentation d’images biomédicales, Yang *et al.* (2017) sont parvenus à obtenir des résultats comparables à ceux de la littérature avec seulement 50 % de la quantité de l’ensemble d’entraînement annoté.

Ainsi, même si l’utilisation d’images du domaine cible semble incontournable pour permettre une analyse de bonne qualité dans ce domaine, plusieurs approches existent pour améliorer le passage d’un modèle d’un domaine à un autre sans pour autant passer de nouveau au travers de l’ensemble du processus d’acquisition et d’annotation. Alléger ce processus permet alors de faciliter l’exploitation de jeux de données plus larges et donc de potentiellement mieux représenter la réalité que l’on souhaite étudier, aboutissant à des modèles plus robustes. Explorer et développer ces méthodes semble une perspective prometteuse pour que le manque de données (annotées ou non) ne soit plus le frein qu’il représente actuellement.

Chapitre 6

Discussion générale et conclusion

6.1 Contributions à la recherche

Ce projet de doctorat avait comme objectif général d’évaluer la capacité des RNCs comme technique de traitement d’images pour la reconnaissance de problèmes phytosanitaires de la vigne.

Dans le **Chapitre 2**, nous avons commencé par constater que les RNCs connaissaient une popularité croissante depuis 2016 dans la communauté de recherche travaillant à la reconnaissance automatique des maladies des cultures. Les travaux publiés faisaient état de capacités d’analyse supérieures à celles des autres techniques utilisées jusqu’alors. Cependant, plusieurs études de notre corpus menaient leurs expérimentations sur des images acquises dans des conditions contrôlées. Ces conditions sont très éloignées de la réalité du champ, ce que le modèle que nous visions doit être capable de gérer. Également, plusieurs études manquaient de rigueur méthodologique, ce qui laissait craindre une surestimation des résultats présentés.

Dès notre premier cas d’étude, présenté dans le **Chapitre 3** et centré sur la reconnaissance du mildiou, nous avons pu éprouver le potentiel des RNCs pour la classification d’images prises en milieu non contrôlé. En entraînement et en inférence, des portions d’images complexes, prises notamment sur différents organes de la vigne mais aussi sur plusieurs niveaux d’avancement et d’intensité de symptômes, ont été présentées au modèle. Lors de l’évaluation, celui-ci a été capable de distinguer les sept classes à 95,48 %. Dans le second cas d’étude, présenté dans le **Chapitre 4** et se focalisant sur la reconnaissance de symptômes de FD, nous nous sommes penchés sur la segmentation, qui procure un rendu plus adapté à un outil de prospection. Plus précisément, un classifieur a tout d’abord été entraîné puis converti en RPC pour l’inférence. Les taux de vrais positifs obtenus en évaluation sur des images venant du même cépage que celui utilisé en entraînement s’élevait à 98,48 % sur les images issues de notre propre acquisition, et à 100 % sur un ensemble réduit d’images provenant d’une source externe. Le maintien d’une haute capacité d’analyse sur des images prises par une personne tierce souligne la robustesse du modèle au type de caméra utilisée pour l’acquisition. Nous avons également constaté les avantages liés au choix d’une taxonomie binaire. Cela a tout d’abord permis de simplifier le processus d’annotation grâce au rassemblement de toutes les images non FD en une seule classe. Sur le terrain, le diagnostic était en effet parfois difficile, même par des experts chevronnés. La taxonomie binaire a ainsi permis de surmonter cette absence de diagnostic. Elle a également permis une gestion simplifiée des cas de cohabitation de maladies, comme celle de FD jumelée à du mildiou, très présente dans nos ensembles de données. Enfin, en utilisant des techniques avancées de visualisation, nous avons pu aboutir à une meilleure compréhension des facteurs influençant les prédictions, permettant ainsi une plus grande maîtrise du modèle. Dans le **Chapitre 5**, nous nous sommes penchés sur la capacité d’application des modèles sur des images à différentes résolutions spatiales. Nous avons pu constater que, dans notre cas, la résolution spatiale la plus élevée n’était pas celle qui apportait les meilleures prédictions et qu’elle conduisait au contraire à une plus grande présence de faux positifs. A la fois lors de l’inférence sur des images de différentes résolutions spatiales et provenant d’une autre plateforme d’acquisition, le

drone, nous avons également pu observer l’impact de la sélection des annotations pour constituer le jeu d’entraînement, même si celles-ci sont issues des mêmes images.

Ainsi, les trois études menées dans ce projet de doctorat ont prouvé que les RNCs permettaient de fournir des modèles d’analyse aux performances robustes. Les RNCs sont donc tout à fait appropriés pour le développement d’outils de prospection viticole. Les études réalisées nous ont également permis de mieux maîtriser les spécificités liées à l’application en contexte agricole. Cependant, au cours de nos expérimentations, nous avons également pu identifier des limites liées à l’utilisation de RNCs pour la reconnaissance de problèmes phytosanitaires de la vigne.

Tout d’abord, nous avons été confrontés à des cas de divergence de données, résultant à une incapacité des modèles à analyser les images concernées. Dans le **Chapitre 4**, cette divergence était liée au cépage, pour lequel l’expression des symptômes de FD est amenée à varier. Le modèle se basait sur des caractéristiques très présentes sur le cépage utilisé en entraînement, mais moins visibles sur d’autres. Dans le **Chapitre 5**, cette divergence était liée à la plateforme d’acquisition – la perche pour l’entraînement et le drone pour l’inférence. La différence de point de vue et de résolution spatiale apportaient des changements trop conséquents pour permettre au modèle de généraliser sur des images provenant de cette seconde source. Dans le **Chapitre 5**, nous avons également pu constater la difficulté de l’amélioration des performances des modèles de RNCs. En théorie, un ensemble d’entraînement composé de plus d’images et d’exemples difficiles à gérer pour le modèle permet d’aboutir à de meilleures performances. Toutefois, malgré des ajouts d’annotations, nous ne sommes pas parvenus à dépasser les performances obtenues dans le **Chapitre 4** sur les images aux dimensions originales. De manière transversale, nous avons également pu éprouver les difficultés pratiques liées à la création de jeux de données pour la reconnaissance de problèmes phytosanitaires. En effet, l’inoculation n’est pas une option pour toutes les maladies, particulièrement si l’échantillonnage a lieu dans une exploitation commerciale. La campagne d’acquisition requiert alors la surveillance d’un territoire conséquent pour trouver les maladies d’intérêt, ainsi que de la réactivité pour l’acquisition, tout en s’intégrant au calendrier d’opérations agricoles. Constituer un ensemble de données représentatif de la problématique étudiée, de la diversité variétale, phénologique et phytosanitaire constitue ainsi un véritable défi.

6.2 Perspectives pour le développement d’outils de prospection automatique des cultures

Des progrès conséquents vers le développement d’outils de prospection automatique ont été réalisés grâce aux techniques d’AP. Cependant, pour qu’un tel outil intègre la routine de gestion phytosanitaire des viticulteurs ou, plus largement, des agriculteurs, des améliorations à la technologie actuelle demeurent nécessaires, notamment en termes d’exactitude, de fiabilité et d’accessibilité.

L’un des principaux leviers pour accroître l’exactitude des diagnostics délivrés est la collection de données plus complètes. Comme nous l’avons mis en avant au long de cette thèse, cette tâche

n'est pas simple à réaliser et bénéficierait grandement de la collaboration de la communauté travaillant sur cette thématique. A la manière de PlantVillage, qui a eu un apport majeur pour les avancées en reconnaissance de maladies des plantes par RNCs, une banque collaborative d'images prises en champ stimulerait assurément la recherche et le développement, tout en apportant une meilleure diversité que les ensembles acquis individuellement. Une banque d'images idéale pourrait également permettre de regrouper des images de différentes résolutions spectrales. Cela pourrait notamment faciliter l'utilisation du proche-infrarouge, un domaine peu exploité pour le moment avec les RNCs malgré les informations sur l'état physiologique de la plante qu'il permet d'obtenir. Ce travail collaboratif présente cependant des difficultés, notamment quant à la régularité des annotations. Pour passer outre les difficultés liées à cette étape tout en bénéficiant de larges ensembles d'images, les approches d'apprentissage semi ou non-supervisées devraient être explorées. Deux techniques, citées précédemment, nous semblent particulièrement prometteuses pour notre thématique : l'apprentissage actif – où les images les plus "utiles" à l'entraînement du modèle sont annotées, et l'apprentissage auto-supervisé – où le modèle apprend des représentations en résolvant des tâches prétextes. Plus généralement, l'ensemble des techniques d'adaptation de domaine seraient pertinentes pour permettre aux modèles développés de généraliser sur des images à la thématique connexe mais aux caractéristiques différentes de celles utilisées en entraînement. Ces techniques sont très diversifiées, allant de l'affinement de l'apprentissage, à l'apprentissage auto-supervisé ou encore à l'utilisation de réseaux adverses génératifs. Leur utilisation permettrait de mieux exploiter des ensembles de données de petites tailles ou non annotés. Une autre approche nous semble également tout à fait appropriée pour les outils de prospection : la détection d'anomalies. Appliquée à la détection de problèmes phytosanitaires, cela reviendrait à repérer toute situation où la plante n'est pas saine. Au lieu d'apprendre au modèle à reconnaître chacun des ravageurs et maladies, l'objectif serait donc de lui apprendre à reconnaître des plantes en santé. Cette approche procurera un diagnostic peu détaillé mais serait plus à même de donner l'alerte pour des situations déviant de la normalité, même si celles-ci n'ont pas été précédemment observées en entraînement. Enfin, travailler avec des données auxiliaires aux images pourraient accroître la capacité des technologies actuelles. Associer les images à des données telles que la date, les conditions météorologiques ou l'historique de contaminations, permettrait de les insérer dans leur contexte d'acquisition, augmentant par exemple la sensibilité de détection du modèle lorsque les conditions météorologiques des jours précédents sont favorables à l'émergence de maladies, ou lorsque l'on entre dans une période connue comme étant de plus grande vulnérabilité. Ainsi, plusieurs avenues peuvent être envisagées pour parvenir à améliorer les modèles de reconnaissance de problèmes phytosanitaires, et s'en saisir permettra assurément d'obtenir de meilleurs diagnostics.

La production de diagnostics de grande qualité est évidemment le premier critère pour que les agriculteurs aient suffisamment confiance en cette technologie pour l'utiliser. Néanmoins, nous pensons également qu'il est important que son fonctionnement présente une certaine transparence. Il s'agit notamment de connaître les caractéristiques sur lesquelles se base le modèle pour faire ses

prédictions. Plusieurs techniques de visualisation existent pour obtenir cette information. Même si l'interprétation de ces visualisations n'est pas toujours aisée, l'utilisation de l'une de ces techniques s'est révélée très enrichissante pour le cas d'étude présenté dans le Chapitre 4. Connaître les forces et faiblesses des modèles, telles que les cas les plus à risque de confusion ou encore les conditions limites d'application, leur confère également plus de transparence. En effet, toutes les confusions n'ont pas le même impact agronomique, certaines sont anodines, car elles nécessitent par exemple le même moyen de lutte, alors que d'autres risquent de conduire à une mauvaise prise en charge et donc à une potentielle aggravation du problème. Là encore, des techniques de visualisation ou encore une succession d'évaluation sur différents ensembles d'images peuvent permettre d'obtenir ces informations. Ajouter une valeur agronomique au poids de l'erreur pourrait également accroître la maîtrise de la sensibilité des modèles. Enfin, la transparence de cette technologie devrait aussi s'étendre à la question de la propriété des données. Les images acquises lors des prospections automatiques constitueront en effet une manne précieuse pour l'amélioration des modèles. Définir clairement à l'avance les données conservées lors de ces opérations ainsi que leur propriétaire sera incontournable pour la transparence de cette technologie.

La prospection automatique étant encore un domaine peu développé, il reste de nombreux aspects à définir, notamment en ce qui concerne les systèmes d'acquisition et de distribution. Pour que cette technologie puisse bénéficier au plus grand nombre, elle devrait être la plus accessible et adaptée possible aux ressources et besoins des agriculteurs. Grâce à l'AP et à ses techniques d'analyse d'images performantes, il n'est pas forcément nécessaire d'utiliser des caméras très avancées pour obtenir des diagnostics précis – du moins, si l'on se cantonne à la détection de symptômes visibles. Des caméras RVB peu onéreuses peuvent donc suffire pour cette application. Quant au système d'acquisition, il pourrait être constitué d'une caméra embarquée sur du matériel agricole pré-existant, évitant ainsi de nouveaux investissements financiers majeurs. De notre avis, le système d'acquisition idéal doit également demander le minimum d'investissement en temps de l'agriculteur. Une acquisition d'images lors des opérations d'entretien des parcelles ou avec une plateforme autonome constitueraient donc les options idéales.

Ainsi, même si les technologies existantes doivent nécessairement être améliorées et que leur mode de fonctionnement et de déploiement restent à préciser, le contexte actuel est assurément des plus enthousiasmants pour concrétiser le développement d'outils de prospection automatique, que ce soit pour la viticulture ou l'agriculture en général.

Références

- AL-Saddik, H., Simon, J.C. et Cointault, F., 2017, Development of spectral disease indices for ‘flavescence dorée’ grapevine disease identification, *Sensors*, volume 17, n° 12.
- Albetis, J., Jacquin, A., Goulard, M., Poilvé, H., Rousseau, J., Clenet, H., Dedieu, G. et Duthoit, S., 2018, On the potentiality of uav multispectral imagery to detect flavescence dorée and grapevine trunk diseases, *Remote Sensing*, volume 11, n° 1, p. 26.
- Ambiaud, E., 2012, Pratiques phytosanitaires dans la viticulture en 2010, Agreste, Ministère de l’Agriculture et de l’Alimentation, 8 p.
- Arrieta, A.B., Díaz-Rodríguez, N., Ser, J.D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. et Herrera, F., 2020, Explainable artificial intelligence (xai) : Concepts, taxonomies, opportunities and challenges toward responsible ai, *Information Fusion*, volume 58, p. 82–115.
- Atabay, H.A., 2017, Deep residual learning for tomato plant leaf disease identification, *Journal of Theoretical and Applied Information Technology*, volume 95, p. 9.
- Atout France, 2020, Infographie, L’Oenotourisme en France en 2016, <http://www.atout-france.fr/content/oenotourisme>.
- Barbedo, J.G.A., 2016, A review on the main challenges in automatic plant disease identification based on visible range images, *Biosystems engineering*, volume 144, p. 52—60.
- Barbedo, J.G.A., 2018a, Factors influencing the use of deep learning for plant disease recognition, *Biosystems Engineering*, volume 172, p. 84–91.
- Barbedo, J.G.A., 2018b, Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification, *Computers and Electronics in Agriculture*, volume 153, p. 46–53.
- Barriault, E., 2012, Vigne : Guide d’implantation, Centre de référence en agriculture et en agroalimentaire du Québec, Québec, 117 p.
- Bassil, K., Vakil, C., Sanborn, M., Cole, D., Kaur, J. et Kerr, K., 2007, Cancer health effects of pesticides : Systematic review, *Canadian family physician Médecin de famille canadien*, volume 53, p. 1704–1711.
- Bengio, Y., 2012, Practical Recommendations for Gradient-Based Training of Deep Architectures, p. 437–478, dans G. Montavon, G.B. Orr et K.R. Müller (éd.), *Neural Networks : Tricks of the Trade : Second Edition*, Springer, Berlin, Heidelberg.

- Blasco, P.R., 2019, 2019 Statistical Report on World Vitiviniculture, International Organisation of Vine and Wine, 23 p.
- Bonfils, J. et Schvester, D., 1960, Les cicadelles (homoptera auchenorhyncha) dans leurs rapports avec la vigne dans le sud- ouest de la france, *Ann Epiphyt*, volume 9, p. 325–336.
- Boulent, J., Beaulieu, M., St-Charles, P.L., Théau, J. et Foucher, S., 2019a, Deep learning for in-field image-based grapevine downy mildew identification, dans J.V. Stafford (éd.), *Precision agriculture '19*, Wageningen Academic Publisher, p. 141–148.
- Boulent, J., Foucher, S., Théau, J. et St-Charles, P.L., 2019b, Convolutional neural networks for the automatic identification of plant diseases, *Frontiers in Plant Science*, volume 10, p. 941.
- Brahimi, M., Arsenovic, M., Laraba, S., Sladojevic, S., Boukhalfa, K. et Moussaoui, A., 2018, Deep Learning for Plant Diseases : Detection and Saliency Map Visualisation, p. 93–117, dans J. Zhou et F. Chen (éd.), *Human and Machine Learning*, Springer International Publishing, Cham.
- Brahimi, M., Boukhalfa, K. et Moussaoui, A., 2017, Deep Learning for Tomato Diseases : Classification and Symptoms Visualization, *Applied Artificial Intelligence*, volume 31, n° 4, p. 299–315.
- Brisco, B., Brown, R., Hirose, T., McNairn, H. et Staenz, K., 1998, Precision agriculture and the role of remote sensing : A review, *Canadian Journal of Remote Sensing*, volume 24, n° 3, p. 315–327.
- Brodeur, J., Boivin, G., Grenier, P. et Gagnon, A., 2013, Impacts des changements climatiques sur le synchronisme entre les ravageurs et leurs ennemis naturels : conséquences sur la lutte biologique en milieu agricole au Québec, *Ouranos*, 124 p.
- Buda, M., Maki, A. et Mazurowski, M.A., 2018, A systematic study of the class imbalance problem in convolutional neural networks, *Neural Networks*, volume 106, p. 249–259.
- Butault, J.P., Delame, N., Jacquet, F. et Zardet, G., 2011, L'utilisation des pesticides en France : état des lieux et perspectives de réduction, p. 7–26, dans *Notes et études socio-économiques*, Centre d'études et de prospective, Ministère de l'agriculture et de l'alimentation.
- Cardona, O.D., 2003, The need for rethinking the concepts of vulnerability and risk from a holistic perspective : A necessary review and criticism for effective risk management., p. 37–51, dans D.H. G. Bankoff G. Frerks (éd.), *Mapping Vulnerability : Disasters, Development and People*, Earthscan Publishers, London, 256 p.
- Carisse, O., 2009, Gestion raisonnée des principales maladies de la vigne au Québec, *Agriculture et Agroalimentaire Canada*, 47 p.
- Carisse, O., Bacon, R.B., Lasnier, J. et McFadden-Smith, W., 2006, Guide d'identification des principales maladies de la vigne, *Agriculture et Agroalimentaire Canada*, 32 p.

- Chalapathy, R. et Chawla, S., 2019, Deep learning for anomaly detection : A survey, arXiv, p. 50, abs/1901.03407.
- Chambre d'Agriculture de Loir-et-Cher, 2015, Référentiel économique du vigneron 2015/2019, 8 p.
- Ching, T., Himmelstein, D.S., Beaulieu-Jones, B.K., Kalinin, A.A., Do, B.T., Way, G.P., Ferrero, E., Agapow, P.M., Zietz, M., Hoffman, M.M., Xie, W., Rosen, G.L., Lengerich, B.J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A.E., Shrikumar, A., Xu, J., Cofer, E.M., Lavender, C.A., Turaga, S.C., Alexandari, A.M., Lu, Z., Harris, D.J., DeCaprio, D., Qi, Y., Kundaje, A., Peng, Y., Wiley, L.K., Segler, M.H.S., Boca, S.M., Swamidass, S.J., Huang, A., Gitter, A. et Greene, C.S., 2018, Opportunities And Obstacles For Deep Learning In Biology And Medicine, *Journal of the Royal Society Interface*, volume 15, n° 141, p. 48.
- Chollet, F., 2017, Deep learning with Python, Manning Publications Co., 384 p.
- Chuche, J. et Thiery, D., 2014, Biology and ecology of the flavescence doree vector scaphoideus titanus : a review, *Agronomy for Sustainable Development*, volume 34, n° 12, p. 381–403.
- Cooper, J. et Dobson, H., 2007, The benefits of pesticides to mankind and the environment, *Crop Protection*, volume 26, n° 9, p. 1337–1348.
- Corfield, D., 2009, Projection and projectability, p. 248, dans J. Quionero-Candela, M. Sugiyama, A. Schwaighofer et N.D. Lawrence (éd.), *Dataset shift in machine learning*, The MIT Press.
- Coulon, T. et Hugueniot, F., 2008, Production intégrée de raisins - De la conception à l'application : état des pratiques et voies d'amélioration, Institut français de la vigne et du vin, Cahiers Itinéraires, volume 17, 36 p.
- Cruz, A., Ampatzidis, Y., Pierro, R., Materazzi, A., Panattoni, A., Bellis, L.D. et Luvisi, A., 2019, Detection of grapevine yellows symptoms in vitis vinifera l. with artificial intelligence, *Computers and Electronics in Agriculture*, volume 157, p. 63–76.
- Cruz, A.C., Luvisi, A., De Bellis, L. et Ampatzidis, Y., 2017, X-FIDO : An Effective Application for Detecting Olive Quick Decline Syndrome with Deep Learning and Data Fusion, *Frontiers in Plant Science*, volume 8, p. 1741.
- Curran, P.J., 1985, Aerial photography for the assessment of crop condition : a review, *Applied Geography*, volume 5, n° 4, p. 347–360.
- Darriet, P., Pons, A., Thibon, C., Schüttler, A., Allamy, L. et Van Leeuwen, C., 2016, Quels impacts attendus du changement climatique sur les composés aromatiques et leurs précurseurs ?, dans 3èmes assises des vins du Sud-Ouest, Institut français de la vigne et du vin, Toulouse, France, p. 19–22.

- DeChant, C., Wiesner-Hanks, T., Chen, S., Stewart, E.L., Yosinski, J., Gore, M.A., Nelson, R.J. et Lipson, H., 2017, Automated Identification of Northern Leaf Blight-Infected Maize Plants from Field Imagery Using Deep Learning, *Phytopathology*, volume 107, n° 11, p. 1426–1432.
- Direction Générale des Entreprises, 2017, Chiffres Clés du Tourisme, édition 2017, Ministère français de l'Economie et des Finances, 6 p.
- Doran, D., Schulz, S. et Besold, T.R., 2018, What does explainable ai really mean ? a new conceptualization of perspectives, *CEUR Workshop Proceedings*, volume 2071, p. 8.
- Fanjul, M.J., 2017, Technical Datasheets In Innovative Solutions To Control Grapevine Trunk Disease And Flavescence Dorée, *WinetWork Project*, 27 p.
- Ferentinos, K.P., 2018, Deep learning models for plant disease detection and diagnosis, *Computers and Electronics in Agriculture*, volume 145, p. 311–318.
- Filippin, L., Jović, J., Cvrković, T., Forte, V., Clair, D., Toševski, I., Boudon-Padieu, E., Borgo, M. et Angelini, E., 2009, Molecular characteristics of phytoplasmas associated with flavescence dorée in clematis and grapevine and preliminary results on the role of *dictyophara europaea* as a vector, *Plant Pathology*, volume 58, n° 5, p. 826–837.
- Firrao, G., Andersen, M., Bertaccini, A., Boudon, E., Bové, J., Daire, X., Davis, R., Fletcher, J., Garnier, M., Gibb, K., Gundersen-Rindal, D., Harrison, N., Hiruki, C., Kirkpatrick, B., Jones, P., Kuske, C., Lee, I., Liefting, L., Marcone, C. et Wang, K., 2004, 'candidatus phytoplasma', a taxon for the wall-less, non-helical prokaryotes that colonize plant phloem and insects, *International Journal of Systematic and Evolutionary Microbiology*, volume 54, p. 1243–1255.
- Food and Agriculture Organization of the United Nation, 2018, FAOSTAT Pesticides Use Dataset, online, accessed in Feb. 2019; <http://www.fao.org/faostat/en/#data/RP>.
- Food and Agriculture Organization of the United Nation, International Plant Protection Convention, 2017, Plant health and food security, pamphlet I7829EN/1/09.17, Accessed online in Feb. 2019.
- Fuentes, A., Yoon, S., Kim, S. et Park, D., 2017, A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition, *Sensors*, volume 17, n° 9, p. 2022.
- Fuentes, A.F., Yoon, S., Lee, J. et Park, D.S., 2018, High-Performance Deep Neural Network-Based Tomato Plant Diseases and Pests Diagnosis System With Refinement Filter Bank, *Frontiers in Plant Science*, volume 9, p. 1162.
- Gill, H. et Garg, H., 2014, Pesticides : Environmental impacts and management strategies, *InTech*, p. 188–230.

- Goodfellow, I.J., Bengio, Y. et Courville, A., 2016, Deep Learning, MIT Press, Cambridge, MA, USA, <http://www.deeplearningbook.org>.
- Gorriz, M., Carlier, A., Faure, E. et Giró i Nieto, X., 2017, Cost-effective active learning for melanoma segmentation, ArXiv, p. 5, abs/1711.09168.
- Gould, S., Fulton, R. et Koller, D., 2009, Decomposing a scene into geometric and semantically consistent regions, dans Proceedings of the 12th International Conference on Computer Vision, ICCV'09, p. 8.
- Goulson, D., 2014, Ecology pesticides linked to bird declines, Nature, volume 511, p. 295–296.
- Hall, A., Lamb, D., Holzapfel, B. et Louis, J., 2002, Optical remote sensing applications in viticulture - a review, Australian Journal of Grape and Wine Research, volume 8, n° 1, p. 36–47.
- He, K., Zhang, X., Ren, S. et Sun, J., 2016, Deep residual learning for image recognition, dans 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas (USA), p. 770–778.
- Hendrycks, D., Mazeika, M., Kadavath, S. et Song, D., 2019, Using self-supervised learning can improve model robustness and uncertainty, dans H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox et R. Garnett (éd.), Advances in Neural Information Processing Systems 32, Curran Associates, Inc., p. 15 663–15 674.
- Hinnewinkel, J.C., 2010, Quand le climat impose une mutation des pratiques : Le retour des terroirs ?, p. 201–229, dans Le changement climatique : Quand le climat nous pousse à changer d'ère, Presses universitaires du Septentrion, Villeneuve d'Ascq.
- Holzinger, A., Biemann, C., Pattichis, C.S. et Kell, D.B., 2017, What do we need to build explainable ai systems for the medical domain ?, abs/1712.09923.
- Hughes, D.P. et Salathé, M., 2015, An open access repository of images on plant health to enable the development of mobile disease diagnostics, p. 13.
- Hutson, M., 2018, Artificial intelligence faces reproducibility crisis, Science, volume 359, n° 6377, p. 725–726.
- Jing, L. et Tian, Y., 2019, Self-supervised visual feature learning with deep neural networks : A survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, *Preprint*.
- Johnson, H. et Robinson, J., 2013, The World Atlas of Wine, Ed. Octopus Publishing Group, London, 400 p.
- Kamilaris, A. et Prenafeta-Boldú, F.X., 2018, Deep learning in agriculture : A survey, Computers and Electronics in Agriculture, volume 147, p. 70–90.

- Kawahara, J. et Hamarneh, G., 2016, Multi-resolution-tract cnn with hybrid pretrained and skin-lesion trained layers, dans L. Wang, E. Adeli, Q. Wang, Y. Shi et H.I. Suk (éd.), *Machine Learning in Medical Imaging*, Springer International Publishing, Cham, p. 164–171.
- Khan, A., Sohail, A., Zahoor, U. et Qureshi, A.S., 2020, A survey of the recent architectures of deep convolutional neural networks, *Artificial Intelligence Review*, volume 53, p. 5455–5516.
- Kim, K.H., Kabir, E. et Ara Jahan, S., 2016, Exposure to pesticides and the associated human health effects, *Science of The Total Environment*, volume 575, p. 525–535.
- Kingma, D.P. et Ba, J., 2015, Adam : A method for stochastic optimization, dans Y. Bengio et Y. LeCun (éd.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, USA, May 7-9, 2015*.
- Knillmann, S. et Liess, M., 2019, Pesticide Effects on Stream Ecosystems, p. 211–214, dans M. Schröter, A. Bonn, S. Klotz, R. Seppelt et C. Baessler (éd.), *Atlas of Ecosystem Services : Drivers, Risks, and Societal Responses*, Springer International Publishing, Cham.
- Knipling, E.B., 1970, Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation, *Remote Sensing of Environment*, volume 1, n° 3, p. 155–159.
- Kokhlikyan, N., Miglani, V., Martin, M., Wang, E., Reynolds, J., Melnikov, A., Lunova, N. et Reblitz-Richardson, O., 2019, Pytorch captum, <https://github.com/pytorch/captum>.
- Krawczyk, B., 2016, Learning from imbalanced data : open challenges and future directions, *Progress in Artificial Intelligence*, volume 5, n° 4, p. 221–232.
- Kumar, L., Schmidt, K., Dury, S. et Skidmore, A., 2001, Imaging spectrometry and vegetation science, p. 111–155, dans F. van der Meer et S. de Jong (éd.), *Imaging spectrometry : basic principles and prospective applications (Remote sensing and digital image processing)*, Kluwer Academic.
- Kumar, N., Byrappa, A. et Gopalkrishna, R., 2020, Soil Fauna and Sustainable Agriculture, p. 211–226, dans A.K. Chakravarthy (éd.), *Innovative Pest Management Approaches for the 21st Century*, Springer International Publishing, Cham.
- La Cité du Vin, 2020, Le vin, patrimoine culturel, universel et vivant, <https://fondation.laciteduvin.com/fr/la-fondation-pour-la-culture-les-civilisations-du-vin>.
- LeCun, Y., Bengio, Y. et Hinton, G., 2015, Deep learning, *Nature*, volume 521, n° 7553, p. 436–444.
- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciampi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B. et Sánchez, C.I., 2017, A survey on deep learning in medical image analysis, *Medical Image Analysis*, volume 42, p. 60–88.

- Liu, B., Zhang, Y., He, D. et Li, Y., 2017a, Identification of Apple Leaf Diseases Based on Deep Convolutional Neural Networks, *Symmetry*, volume 10, n° 1, p. 11.
- Liu, P., Zhang, H. et Eom, K.B., 2017b, Active deep learning for classification of hyperspectral images, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, volume 10, n° 2, p. 712–724.
- Lucas, P., 2007, Le concept de la protection intégrée des cultures, *Innovations Agronomiques*, volume 1, p. 15–21.
- Luginbühl, Y., 2005, Paysages viticoles, p. 15–20, dans *Les Paysages Culturels Viticoles*, Conseil International des Monuments et des Sites, Paris.
- Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., Villa, P., Stroppiana, D., Boschetti, M., Goulart, L., Davis, C. et Dandekar, A., 2015, Advanced methods of plant disease detection. a review, *Agronomy for Sustainable Development*, volume 35, n° 1, p. 1–25.
- McInnes, L., Healy, J., Saul, N. et Grossberger, L., 2018, Umap : Uniform manifold approximation and projection, *The Journal of Open Source Software*, volume 3, n° 29, p. 861.
- Ministère de la Transition Ecologique et Solidaire et Ministère de l'Agriculture et de l'Alimentation, 2018, Ecophyto, Note de suivi 2017, République française, 36 p.
- Ministère de l'Environnement, du Développement durable et de la Lutte contre les changements climatiques, 2015, Stratégie québécoise sur les pesticides 2015-2018, Gouvernement du Québec, 28 p.
- Ministère de l'Écologie et du Développement durable et Ministère de l'Agriculture et de l'Alimentation, 2015, Plan Ecophyto II+, République française, 66 p.
- Ministère français de l'agriculture, d.l.e.d.l.f., 2013, Arrêté du 19 décembre 2013 relatif à la lutte contre la flavescence dorée de la vigne et contre son agent vecteur, online, accessed in Feb. 2020 ; <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000028409891>.
- Mohanty, S.P., Hughes, D.P. et Salathé, M., 2016, Using Deep Learning for Image-Based Plant Disease Detection, *Frontiers in Plant Science*, volume 7, p. 1419.
- Mulla, D.J., 2013, Twenty five years of remote sensing in precision agriculture : Key advances and remaining knowledge gaps, *Biosystems Engineering*, volume 114, n° 4, p. 358–371, special Issue : Sensing Technologies for Sustainable Agriculture.
- Oppenheim, D. et Shani, G., 2017, Potato Disease Classification Using Convolution Neural Networks, *Advances in Animal Biosciences*, volume 8, n° 02, p. 244–249.

- Organisation des Nations unies pour l'éducation, l.s.e.l.c., 2020, Liste du patrimoine mondial, recherche pour le mot clef "vigne", <https://whc.unesco.org/fr/list/?search=vigne&order=country>.
- Otálora, S., Perdomo, O., González, F. et Müller, H., 2017, Training Deep Convolutional Neural Networks with Active Learning for Exudate Classification in Eye Fundus Images, p. 146 – 154.
- Outreville, J.F., 2009, Wine Production in Québec, *Bacchus Goes Green*, 9, 10, 11 juillet 2009, Dijon, France, p. 15.
- Peng, R.D., 2011, Reproducible research in computational science, *Science*, volume 334, n° 6060, p. 1226–1227.
- Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J. et Johannes, A., 2018, Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild, *Computers and Electronics in Agriculture*, volume 161, p. 280 – 290.
- Pimentel, D. et Greiner, A., 1997, Environmental and socioeconomic costs of pesticide use, p. 51–78, dans J. Wiley et Sons (éd.), *Techniques for Reducing Pesticide Use : Economic and Environmental Benefits*, Chichester.
- Powers, D.M.W., 2011, Evaluation : From precision, recall and f-measure to roc., informedness, markedness & correlation, *Journal of Machine Learning Technologies*, volume 2, n° 1, p. 37–63.
- Qin, Z., Yu, F., Liu, C. et Chen, X., 2018, How convolutional neural network see the world - A survey of convolutional neural network visualization methods, *Mathematical Foundations of Computing*, volume 1, p. 149–180.
- Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J. et Hughes, D.P., 2017, Deep Learning for Image-Based Cassava Disease Detection, *Frontiers in Plant Science*, volume 8, p. 1852.
- Ramcharan, A., McCloskey, P., Baranowski, K., Mbilinyi, N., Mrisho, L., Ndalawwa, M., Legg, J. et Hughes, D.P., 2019, A mobile-based deep learning model for cassava disease diagnosis, *Frontiers in Plant Science*, volume 10, p. 272.
- Reynolds, A., 2017, The Grapevine, Viticulture, and Winemaking : A Brief Introduction, p. 3–29, dans B. Meng, G.P. Martelli, D.A. Golino et M. Fuchs (éd.), *Grapevine Viruses : Molecular Biology, Diagnostics and Management*, Springer International Publishing, Cham, 709 p.
- Risebrough, R.W., 1986, Pesticides and Bird Populations, p. 397–427, dans R.F. Johnston (éd.), *Current Ornithology : Volume 3*, Springer US, Boston (USA).

- Rong, G., Mendez, A., Assi], E.B., Zhao, B. et Sawan, M., 2020, Artificial intelligence in healthcare : Review and prediction case studies, *Engineering*, volume 6, n° 3, p. 291–301.
- Rosenzweig, C., Iglesias, A., Yang, X.B., Epstein, P.R. et Chivian, E., 2001, Climate change and extreme weather events : Implications for food production, plant diseases, and pests, *Glob. Change Human Health*, volume 2, p. 90–104.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C. et Fei-Fei, L., 2015, Imagenet large scale visual recognition challenge, *Int. J. Comput. Vision*, volume 115, n° 3, p. 211–252.
- Russell, S.J. et Norvig, P., 1995, *Artificial Intelligence A Modern Approach*, Prentice Hall, Englewood Cliffs, 946 p.
- Sabottke, C.F. et Spieler, B.M., 2020, The effect of image resolution on deep learning in radiography, *Radiology : Artificial Intelligence*, volume 2, n° 1, p. 7.
- Saguez, J., 2017, Impact des changements climatiques et mesures d’adaptations pour les ravageurs présents et potentiels en grandes cultures au Québec, CEROM, 71 p.
- Sanchez-Bayo, F. et Goka, K., 2014, Pesticide residues and bees – a risk assessment, *Plos One*, volume 9, n° 4, p. 1–16.
- Sankaran, S., Mishra, A., Ehsani, R. et Davis, C., 2010, A review of advanced techniques for detecting plant diseases, *Computers and Electronics in Agriculture*, volume 72, n° 1, p. 1–13.
- Schimmelpfennig, D., 2016, Farm profits and adoption of precision agriculture, , n° 1477-2016-121190, p. 46.
- Schramowski, P., Stammer, W., Teso, S., Brugger, A., Shao, X., Luigs, H.G., Mahlein, A.K. et Kersting, K., 2020, Right for the wrong scientific reasons : Revising deep networks by interacting with their explanations, *Preprint*.
- Seguin, B., 2007, Le réchauffement climatique et ses conséquences pour la viticulture, dans Réchauffement climatique, quels impacts probables sur les vignobles?, 28-30 mars 2007, Dijon, France.
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. et Batra, D., 2019, Grad-cam : Visual explanations from deep networks via gradient-based localization, *International Journal of Computer Vision*, volume 128, n° 2, p. 336–359.
- Settles, B., 2009, *Active Learning Literature Survey*, University of Wisconsin–Madison, Computer Sciences Technical Report 1648, 47 p.

- Shelhamer, E., Long, J. et Darrell, T., 2017, Fully convolutional networks for semantic segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 39, n° 4, p. 640–651.
- Simonyan, K. et Zisserman, A., 2015, Very deep convolutional networks for large-scale image recognition, dans *International Conference on Learning Representations (ICLR)*, Mai 2015, San Diego (USA), p. 14.
- Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D. et Stefanovic, D., 2016, Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification, *Computational Intelligence and Neuroscience*, volume 2016, p. 1–11.
- Smith, L.N., 2018, A disciplined approach to neural network hyper-parameters : Part 1 – learning rate, batch size, momentum, and weight decay, *ArXiv*, p. 21, abs/1803.09820.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. et Salakhutdinov, R., 2014, Dropout : A Simple Way to Prevent Neural Networks from Overfitting, p. 30.
- Storkey, A., 2009, When training and test sets are different : Characterizing learning transfer, dans J. Quionero-Candela, M. Sugiyama, A. Schwaighofer et N.D. Lawrence (éd.), *Dataset shift in machine learning*, The MIT Press.
- Sun, S., 2013, A survey of multi-view machine learning, *Neural Computing and Applications*, volume 23, p. 2031–2038.
- Sun, Y., Tzeng, E., Darrell, T. et Efros, A.A., 2019, Unsupervised domain adaptation through self-supervision, *ArXiv*, abs/1909.11825.
- Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B. et Liang, J., 2016, Convolutional neural networks for medical image analysis : Full training or fine tuning?, *IEEE Transactions on Medical Imaging*, volume 35, n° 5, p. 1299–1312.
- Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C. et Liu, C., 2018, A survey on deep transfer learning, dans V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis et I. Maglogiannis (éd.), *Artificial Neural Networks and Machine Learning – ICANN 2018*, Springer International Publishing, Cham, p. 270–279.
- Too, E.C., Yujian, L., Njuki, S. et Yingchun, L., 2018, A comparative study of fine-tuning deep learning models for plant disease identification, *Computers and Electronics in Agriculture*, volume 161, p. 272–279.
- Touvron, H., Vedaldi, A., Douze, M. et Jegou, H., 2019, Fixing the train-test resolution discrepancy, p. 8252–8262, dans H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox et R. Garnett (éd.), *Advances in Neural Information Processing*, Curran Associates, Inc.

- Transports Canada, 2020, Utiliser votre drone de façon sécuritaire et légale, <https://tc.canada.ca/fr/aviation/securite-drones/utiliser-votre-drone-facon-securitaire-legale>, accessed : 22.09.2020.
- Van Leeuwen, C. et Darriet, P., 2016, Le changement climatique en viticulture : les leviers d’adaptation au vignoble, dans 3èmes assises des vins du Sud-Ouest, Toulouse, France, Institut français de la vigne et du vin, p. 19–22.
- Wang, G., Sun, Y. et Wang, J., 2017, Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning, *Computational Intelligence and Neuroscience*, volume 2017, p. 1–8.
- Wang, M. et Deng, W., 2018, Deep visual domain adaptation : A survey, *Neurocomputing*, volume 312, p. 135–153.
- Weisenburger, D.D., 1993, Human health effects of agrichemical use, *Human Pathology*, volume 24, n° 6, p. 571–576.
- Wiesner-Hanks, T., Stewart, E.L., Kaczmar, N., DeChant, C., Wu, H., Nelson, R.J., Lipson, H. et Gore, M.A., 2018, Image set for deep learning : field images of maize annotated with disease symptoms, *BMC Research Notes*, volume 11 :440, p. 3.
- Wilson, C. et Tisdell, C., 2001, Why farmers continue to use pesticides despite environmental, health and sustainability costs, *Ecological Economics*, volume 39, n° 3, p. 449–462.
- Yang, L., Zhang, Y., Chen, J., Zhang, S. et Chen, D.Z., 2017, Suggestive annotation : A deep active learning framework for biomedical image segmentation, dans M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D.L. Collins et S. Duchesne (éd.), *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, Springer International Publishing, Cham, p. 399–407.
- Zeiler, M.D. et Fergus, R., 2014, Visualizing and understanding convolutional networks, dans D. Fleet, T. Pajdla, B. Schiele et T. Tuytelaars (éd.), *Computer Vision – ECCV 2014*, Springer International Publishing, Cham, p. 818–833.
- Zhang, K., Wu, Q., Liu, A. et Meng, X., 2018a, Can Deep Learning Identify Tomato Leaf Disease ?, *Advances in Multimedia*, volume 2018, p. 1–10.
- Zhang, N., Wang, M. et Wang, N., 2002, Precision agriculture—a worldwide overview, *Computers and Electronics in Agriculture*, volume 36, n° 2, p. 113–132.
- Zhang, S., Huang, W. et Zhang, C., 2018b, Three-channel convolutional neural networks for vegetable leaf disease recognition, *Cognitive Systems Research*, volume 53, p. 31–41.

Annexes

Annexe A : Description des données

La totalité des images utilisées pour l’entraînement des RNCs, ainsi que la grande majorité des images utilisées pour l’évaluation des modèles, ont été acquises dans le cadre de ce projet. Dans cette annexe, une description de la phase d’acquisition, des données collectées et de leur organisation sont décrites.

Acquisition des données

Types d’acquisition

Justification du choix de plateformes Initialement, ce projet visait à utiliser majoritairement des images acquises par drone. Cependant, plusieurs contraintes se sont posées : (1) La disponibilité du matériel et du pilote. (2) La nécessité d’obtenir des autorisations de vol, ce qui implique des démarches administratives, donc une prévision du vol plusieurs semaines à l’avance, et qui implique également l’exclusion de zones au survol interdit, notamment à cause de la présence d’habitations ou de couloirs réservés aux exercices militaires. (3) Le besoin d’une UHRS, ce qui implique des vols manuels à basse altitude, donc de longs temps de vol et des changements de batterie fréquents. Ces contraintes étaient peu compatibles avec les réalités biologiques et avec les besoins en données associés à l’utilisation de techniques d’AP. En effet, il n’était pas possible de déterminer à l’avance quelles parcelles allaient être touchées par le mildiou et la FD. Il était donc difficile de gérer le volet administratif associé aux autorisations de vol. Choisir cette option aurait conduit à visiter moins de parcelles. Le jeu de données final aurait donc été moins diversifié en termes de cépages, d’expressivité des symptômes et de conditions météorologiques. Pour pouvoir être réactifs dans l’acquisition et pouvoir visiter un nombre élevé de parcelles, une plateforme additionnelle au drone a été choisie : la perche. Cela nous permettait de nous défaire des contraintes précédemment citées, de nous assurer de l’obtention d’un nombre d’images suffisant tout en faisant l’acquisition d’images avec un point de vue en hauteur, comme l’aurait fait un drone. Des images ont également parfois été prises avec un appareil photo tenu à la main, notamment pendant les phases de repérage de parcelles.

Justification du choix de capteurs Initialement, il était prévu d’évaluer l’apport d’images possédant un canal proche-infrarouge par à celles ne possédant que des canaux du visible. Une caméra RVB-PIR-RedEdge — en l’occurrence la MicaSense Red-Edge — devait être utilisée. Des tests d’acquisition avec cette caméra ont été réalisés en amont du terrain. Durant ces tests, il a été constaté que les images de chaque canal, acquises par des capteurs différents, étaient difficiles à réaligner. Cette difficulté était principalement due à la proximité entre le feuillage et les capteurs. En effet, cette caméra n’a pas été développée pour une telle configuration. Plusieurs corrections ont été entreprises pour essayer d’aligner les canaux au moyen de la librairie OpenCV. Les résultats peu

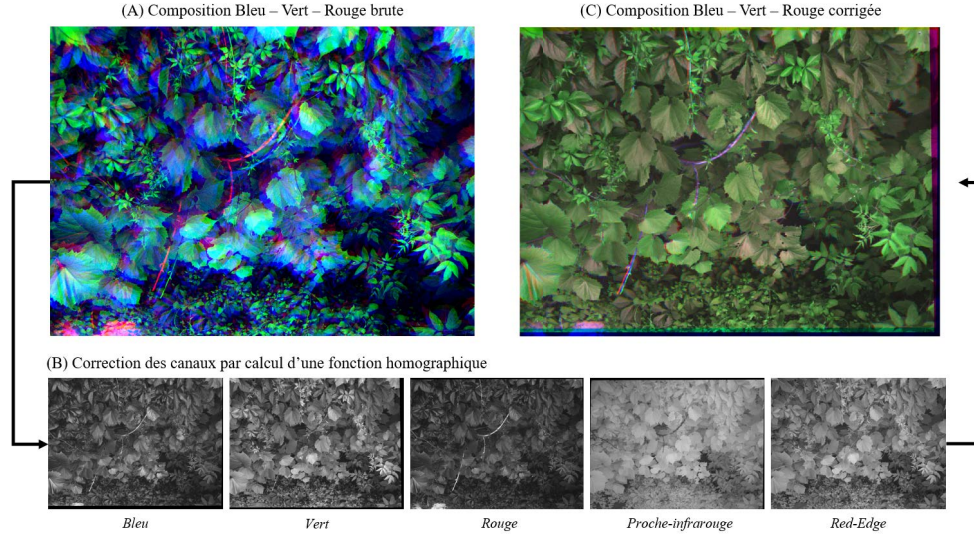


FIGURE 6.1 – Résultat de l’alignement des canaux par calcul d’une fonction homographique sur une image acquise au moyen de la Micasense Red-Edge.

probants (Figure 6.1) nous ont conduits à mettre de côté la caméra MicaSense. Les acquisitions ont donc été exclusivement réalisées avec des caméras RVB.

Description du montage perche Dans ce montage, une perche d’environ 2m a été équipée d’un support pour appareil photo à l’une de ses extrémités. L’appareil photo y était vissé et un déclencheur filaire y était associé par USB, permettant de prendre les photos à distance. Une photographie du montage se trouve en Figure 6.2. L’appareil photo utilisé sur la perche était un Sony de 24,3 mégapixels. Du fait de la structure changeante des vignobles (écart inter-rangs, hauteur des rangs, type de taille), il n’était pas possible de définir des paramètres d’acquisition fixes. La distance entre la caméra et le feuillage ainsi que l’angle d’acquisition varient donc d’une photo à l’autre. Ces paramètres changeaient également selon la localisation des symptômes d’intérêt. De ce fait, les photographies possèdent des visées et résolutions spatiales différentes. Quelques exemples de photos prises avec la perche sont regroupés en Figure 6.3.

Description du montage drone Les vols de drone ont été réalisés par l’entreprise Micro-drones – partenaire du projet. Le drone utilisé était un md4-1000 sur lequel était monté un appareil photo Sony de 42 mégapixels (Figure 6.4). La volonté d’obtenir des images dont la résolution spatiale était d’au moins 1mm a conditionné les paramètres de vol. Pour obtenir une telle résolution spatiale, il était nécessaire de voler à environ 10m d’altitude. Cette hauteur n’était pas supportée par le logiciel de pilotage automatique. Les vols ont donc été effectués en manuel, avec une altitude variant en 8m et 12m et un déclenchement de l’appareil photo toutes les deux secondes. Les photographies ont été prises avec une visée nadir. Quelques exemples de ces photographies prises par drone sont regroupés en Figure 6.5.



FIGURE 6.2 – Perche en cours d'utilisation.



FIGURE 6.3 – Exemples de photos prises avec la perche.



FIGURE 6.4 – Exemples de photos prises avec la perche.



FIGURE 6.5 – Exemples de photos prises avec le drone.

Lieux d'acquisition et partenaires

Les acquisitions se sont déroulées au Canada et en France (Figure 6.6). Au Canada, elles ont eu lieu dans la ferme expérimentale d'Agriculture et Agroalimentaire Canada, située à Frelighsburg, en Montérégie, au Québec. Ces acquisitions ont avant tout servies à constituer le jeu de données orienté mildiou. En France, les acquisitions ont eu lieu dans les régions viticoles de Cognac et du Languedoc-Roussillon, autour de Limoux et de Carcassonne. Ces acquisitions avaient pour objectif la constitution du jeu de données orienté FD. Cependant, l'été 2018 a été très marqué par les contaminations en mildiou dans les deux régions viticoles d'intérêt. Le mildiou est donc également très présent dans les photos acquises dans cette campagne.

Description du site de Montérégie La ferme expérimentale d'Agriculture et Agroalimentaire Canada située à Frelighsburg est constituée d'une centaine d'hectares sur lesquels se trouvent

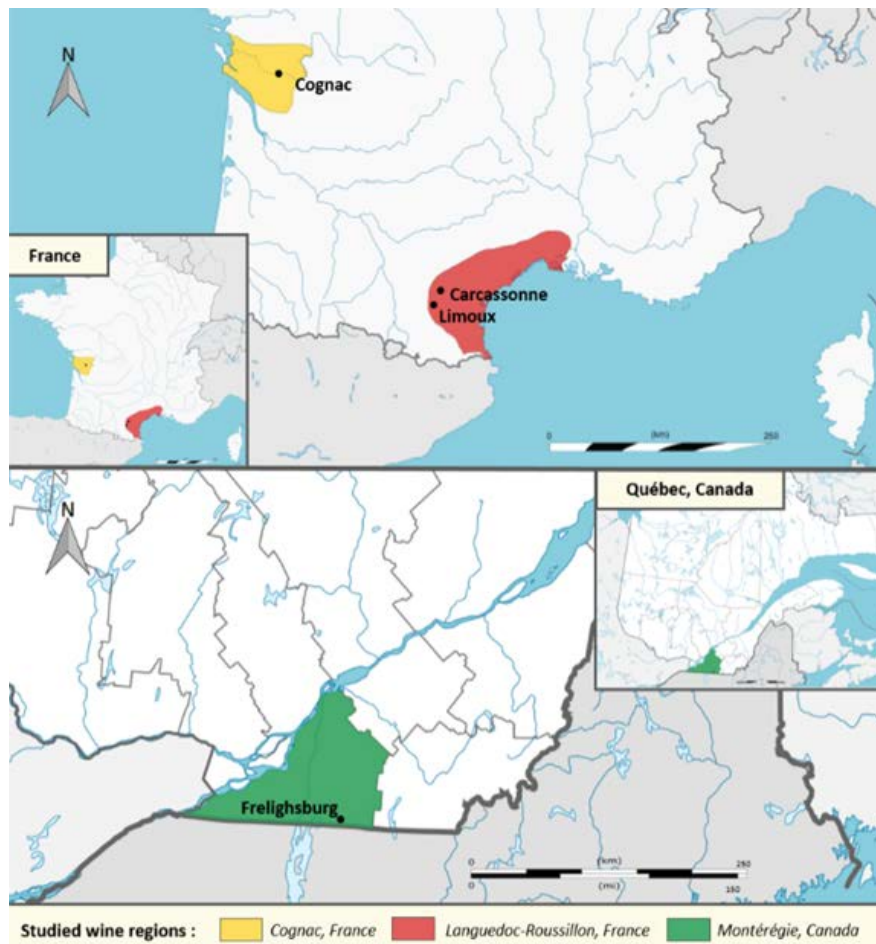


FIGURE 6.6 – Localisation des régions d’acquisition en France et au Canada.

notamment des cultures de petits fruits, des vergers ainsi que des vignes. Deux parcelles sont plantées en vignes (Figure 6.7). La première parcelle (G18) correspond à un carré latin de 12 sections d'une surface totale de 0.23 hectare. Elle est plantée en Chancellor -- un cépage rouge peu exploité en viticulture commerciale mais intéressant de par sa sensibilité à l'oïdium -- depuis 2003. La seconde parcelle (G22) est constituée de 4 blocs plantée de 3 cépages : Gesenheim, Frontenac et Chancellor, pour une surface totale de 0.52 hectare. Des acquisitions avec un appareil photo tenu à la main s'y sont déroulées en juin, juillet et septembre 2017 et des acquisitions avec la perche s'y sont déroulées en juillet 2018. Sur place, Annie Lefebvre, technicienne d'Agriculture et Agroalimentaire Canada, nous a formées à la reconnaissance du mildiou, de l'oïdium, du scarabée japonais et de l'anthracnose. Les problèmes phytosanitaires qui y ont été majoritairement observés sont les suivants : mildiou (sous toutes ses formes), dommages de scarabées japonais et carences en manganèse.



FIGURE 6.7 – Localisation des deux parcelles plantées en vignes (G18 et G22) à la ferme expérimentale d'Agriculture et Agroalimentaire Canada (AAC) à Frelighsburg, d'après une carte fournie par AAC.

Description des sites de Cognac Les acquisitions dans la région de Cognac ont été réalisées dans les vignobles de la maison de Cognac Rémy Martin ainsi que chez des viticulteurs associés. Les acquisitions ont eu lieu sur les communes de Cognac, de Boutiers-Saint-Trojan, de Juillac-le-Coq, de Gensac-la-Pallue, de Saint-Preuil, d'Angeac-Charent et d'Ozillac. 18 parcelles ont été visitées. Le cépage majoritaire était l'Ugni Blanc. Quelques parcelles étaient plantées en Colombard, en Merlot et en Cabernet. Les acquisitions par perche se sont déroulées du 20 au 23 août 2018 et du 10 au 11 septembre 2018. Les acquisitions par drone ont eu lieu le 10 et le 11 septembre 2018 à Ozillac. La parcelle survolée contenait à la fois un cépage blanc (Ugni-Blanc) pour lequel six rangs ont été survolés, et deux cépages rouges (Merlot et Cabernet) pour lesquels trois rangs de chaque ont été

survolés. La localisation des pieds contaminés par la FD a été enregistrée au moyen de l'application portable GPS GéoPaparazzi. Sur place, des salariés du Service Maître de Chai de la maison Rémy Martin nous ont guidés dans les prospections. M. et Mme Branchaud, propriétaires de la parcelle sur laquelle a été réalisé le vol par drone, ont également réalisé une prospection complète de leur parcelle avec nous. De nombreux cas d'esca et de carences en magnésium ont été observés. Le mildiou était également très présent. Quelques cas d'érinose ont été relevés. Plusieurs jaunissements de feuillage pouvant être confondus avec de la FD sur cépages blancs ont également été enregistrés.

Description des sites du Languedoc-Roussillon Les acquisitions dans la région du Languedoc-Roussillon ont été réalisées en partenariat avec une doctorante de l'Ecole de Purpan et de l'entreprise Terranis. Une partie des parcelles visitées avait été sélectionnées lors des saisons précédentes par les équipes de l'École de Purpan et de Terranis dans le cadre d'un autre projet de recherche sur la FD. Le reste des parcelles a été trouvé lors des prospections réalisées au côté d'une chargée de mission de la Chambre d'Agriculture de l'Aude et d'un inspecteur de la Fédération Départementale de Défense contre les Organismes Nuisibles de l'Aude. Ces parcelles étaient situées autour des communes de Carcassonne et de Limoux, plus précisément sur les communes de Barbaira, de Berriac, de La Digne-d'Aval, de Magrie, de Roquetaillade, de Peyriac-Minervois et de Caunes. Sur la vingtaine de parcelles visitées, 12 ont été retenues pour servir à des acquisitions perche et 2 ont été retenues pour les acquisitions par drone. Les acquisitions par perche ont eu lieu du 27 août au 5 septembre 2018. Les vols ont eu lieu les 4 et 5 septembre 2018 à la Digne-d'Aval (cépage : Chardonnay) et à Barbaira (cépages rouges non identifiés). Sur les deux sites, une dizaine de rangs ont été survolés. Les zones survolées ont été complètement prospectées et la localisation des pieds contaminés par la FD a été enregistrée au moyen d'un DGPS. Les photos prises sur cépages blancs étaient pour la plupart du Chardonnay. Quelques photos correspondent à du Chenin. Les photos prises sur cépages rouges correspondent entre autres à du Grenache, du Caladoc et à du Mourvèdre. Quelques cépages n'ont cependant pas pu être identifiés. En plus de la FD, du mildiou, de l'esca, des carences en magnésium, de la phytotoxicité ont été observés.

Données collectées

Récapitulatif quantitatif

Par localisation La Table 6.1 présente un récapitulatif du nombre d'images prises selon leur localisation et par type d'acquisition. Les acquisitions en Montérégie en 2017 n'ayant pas été réalisées spécifiquement pour ce projet, elles ne correspondent qu'à des photos prises à la main. Au total, toutes plateformes confondues, ce sont 8 595 images qui ont été acquises — 60 % via la perche, 31 % à la main et 9 % par drone.

TABLEAU 6.1 – Nombre d’images selon leur localisation et par type d’acquisition.

Localisation – Année	Perche	À la main	Drone
Montérégie – 2017	0	1912	0
Montérégie – 2018	475	0	0
Cognac – 2018	1815	0	362
Languedoc-Roussillon – 2018	2864	742	425
Total	5154	2654	787

Par couleur de cépage La Table 6.2 présente un récapitulatif du nombre d’images prises selon la couleur de cépage et par type d’acquisition. Le nombre d’images prises à la perche est équilibré entre rouge et blanc. Pour les photos prises à la main et par drone, la proportion d’images est plus importante du côté des cépages rouges. Cette différence est notamment due à la difficulté de trouver des parcelles de cépages blancs contaminées par la FD mais aussi au fait que les acquisitions réalisées en Montérégie ne concernaient que des cépages rouges. Ce déséquilibre n’a pas de conséquence pour l’identification du mildiou puisqu’il n’y a pas de variation de symptômes d’une couleur de cépage à l’autre. Pour la FD, même si le jeu de données sera plus réduit en blanc qu’en rouge, il devrait être de taille suffisante pour entraîner des RNCs.

TABLEAU 6.2 – Nombre d’images selon la couleur de cépage et par type d’acquisition. Les photos prises en Montérégie en 2017 n’ont pas pu être incluses à ce tableau de synthèse.

Couleur de cépage	Perche	À la main	Drone
Blanc	2538	265	292
Rouge	2616	2389	495
Total	5154	2654	787

Problèmes phytosanitaires rencontrés

L’expression des symptômes d’un problème phytosanitaire donné n’est pas quelque chose de complètement fixe. En effet, selon les cépages, selon le stade de développement de la vigne et du problème phytosanitaire mais aussi selon le niveau de contamination, l’expression des symptômes va varier. Cette section vise à décrire les situations phytosanitaires rencontrées, des plus emblématiques aux cas plus particuliers. L’identification a été faite selon nos capacités d’expertise. Nous décrivons les deux maladies qui sont ciblées plus spécifiquement dans ce projet : le mildiou et la FD, mais aussi l’ensemble des problèmes phytosanitaires rencontrés lors des prospections.

Mildiou Des vignes contaminées par le mildiou ont été observées à la fois en Montérégie et dans les régions de Cognac et du Languedoc-Roussillon. Plusieurs stades de développement et niveaux d’intensité ont été observés. En Montérégie, les premiers stades de contamination ont pu être observés sur feuilles, tiges et inflorescences (Figure 6.8, A). Ces derniers correspondent aux contaminations

des oospores qui germent dans le sol au printemps en présence d'eau libre. Ce sont à partir de ces contaminations primaires que le mildiou va se répandre via des zoospores. L'inflorescence contaminée est d'abord recouverte par un duvet blanc et prend la forme d'une crosse (Figure 6.8, B et C). Elle finit par dessécher et tomber. Sur les feuilles, les premiers symptômes apparaissent sous forme de taches jaunes relativement circulaires appelées taches d'huile (Figure 6.9 A). Progressivement, ces taches vont sécher, prenant des teintes plus foncées jusqu'à complètement brunir (Figure 6.9, B et C).



FIGURE 6.8 – Expression du mildiou sur inflorescences et tiges. (A) Rameau sain (gauche) et rameau contaminé par le mildiou, stade d'infection primaire (droite). (B) Inflorescence contaminée par le mildiou, stade des sporulations. (C) Inflorescence contaminée par le mildiou, stade avancé des sporulations.



FIGURE 6.9 – Évolution des symptômes foliaires de mildiou : (A) Taches d'huile. (B) Symptômes avancés : les taches tendent du jaune foncé au brun. (C) Symptômes très avancés : les taches sont presque complètement brunes.

La forme de ces taches foliaire peut varier, notamment selon l'intensité de la contamination mais aussi selon la quantité de précipitations ayant conduit à la contamination (le mildiou ayant besoin d'eau pour se développer). Plusieurs faciès différents de l'expression circulaire emblématique ont été observés : des taches sectionnelles présentes sporadiquement (Figure 6.10, A), des taches étendues (Figure 6.10, B), qui peuvent provoquer un repliement foliaire chez les jeunes feuilles ainsi que des taches suivant les nervures (Figure 6.10, C).

Dans certains cas, une combinaison de problèmes phytosanitaires, et donc de symptômes, a pu être observée, notamment avec : une carence (Figure 6.11, A), de l'érinose (Figure 6.11, B) et des



FIGURE 6.10 – Diversité des expressions du mildiou : (A) Taches sectionnelles. (B) Taches étendues. (C) Taches suivant les nervures.

embruns de produits de traitements (Figure 6.11, C). Le cas de la combinaison entre le mildiou et la FD est décrit dans le paragraphe suivant. Hormis ce dernier cas, les associations entre le mildiou et un autre problème phytosanitaire ont été assez peu observés.



FIGURE 6.11 – Association du mildiou avec d'autres problèmes phytosanitaires : (A) Mildiou et carence. (B) Mildiou et l'ériose. (C) Mildiou et embruns de produits de traitement.

Flavescence dorée La FD a surtout été observée dans le Languedoc-Roussillon. Les vignobles de la région de Cognac sont également touchés, mais dans une proportion moindre. Il était donc beaucoup plus difficile de trouver des pieds contaminés. Les premiers symptômes de FD apparaissent l'année suivant la contamination, à la fin du mois de mai et au début du mois de juin. On peut constater une croissance ralentie, des feuilles qui s'enroulent et se rigidifient ainsi que des entre-nœuds raccourcis. Le débourrement est retardé et on peut constater un flétrissement des inflorescences et un port pleureur du pied. A la véraison, les limbes des feuilles deviennent cassants et changent de couleur. Ce sont les symptômes foliaires de changement de couleur et d'enroulement que nous avons cherché à observer. En effet, ils sont les plus faciles à observer par imagerie. Le fait que le rameau n'aôte pas est également un symptôme très caractéristique de la maladie, mais est beaucoup plus difficile à observer. En effet, pour le voir, il fallait souvent ouvrir le feuillage. Sur les cépages blancs, un jaunissement du feuillage va pouvoir être observé (Figure 6.12). Sur les cépages rouges, le feuillage rougit (Figure 6.13). En pratique, une grande variabilité va pouvoir être observée d'un

cépage à l'autre. Certains vont surtout exprimer de l'enroulement (Figure 6.12, A et B, Figure 6.13, A), d'autres le changement de couleur (Figure 6.12, C, Figure 6.13, A et C) et d'autres vont présenter une coloration principalement sur les nervures (Figure 6.12, D, Figure 6.13, B et D).

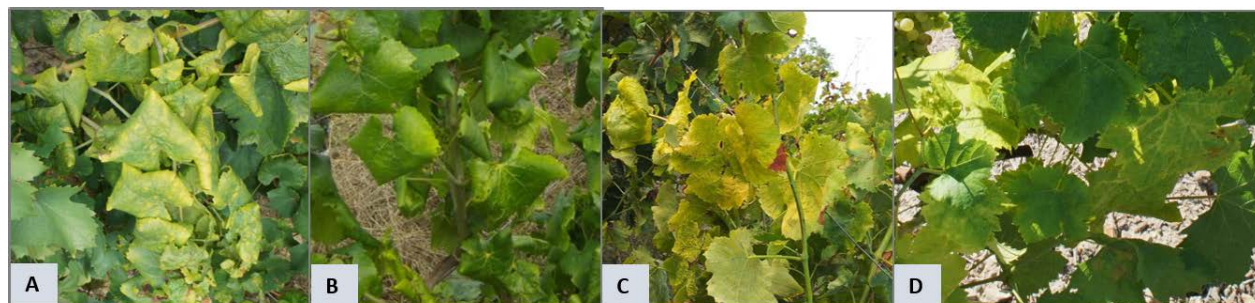


FIGURE 6.12 – Flavescence dorée sur cépages blancs : (A),(B) Chardonnay. (C) Chenin. (D) Ugni-Blanc.



FIGURE 6.13 – Flavescence dorée sur cépages rouges : (A) Grenache. (B) Syrah. (C) Cabernet. (D) Non identifié.

L'année 2018 a été très particulière sur le plan phytosanitaire pour les vignobles français. Beaucoup d'entre eux ont été touchés par le mildiou. Cette maladie apparaissant avant la FD, une association de symptômes a été constatée dans de très nombreux cas (Figure 6.14, A). Des expressions peu courantes de FD pourraient également s'expliquer par des associations (Figure 6.14, B).

D'autres dommages liés à la cicadelle ont pu être observés (Figure 6.15).

Carences Plusieurs carences ont été observées à la fois sur cépages blancs et rouges. Cependant, dans les premiers stades d'expression, identifier assurément une carence n'est pas simple. Plusieurs images de début d'expression de carences restent non identifiées. Entre autres, nous avons pu observer : des carences en magnésium (Figure 6.16, A et B), en manganèse (Figure 6.16, C), en potassium (Figure 6.16, D et E) ou encore en fer (Figure 6.16, F).

Maladies du bois Deux maladies du bois ont été observées : l'esca (Figure 6.17) et le Black Dead Arm (BDA, Figure 6.18). Selon la couleur du cépage, ces deux maladies s'expriment différemment.

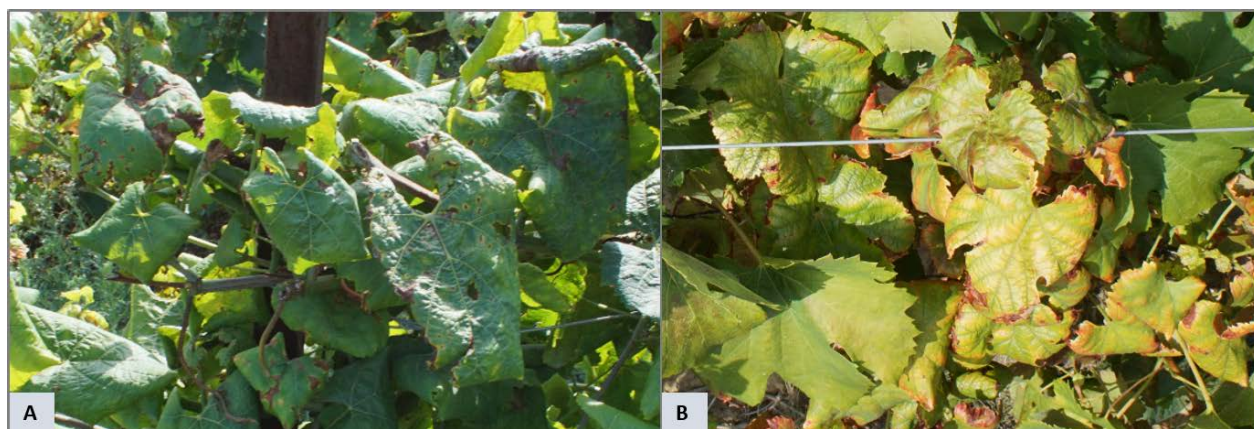


FIGURE 6.14 – Association de symptômes de flavescence dorée avec un autre problème phytosanitaire : (A) Flavescence et mildiou. (B) Flavescence et carence.

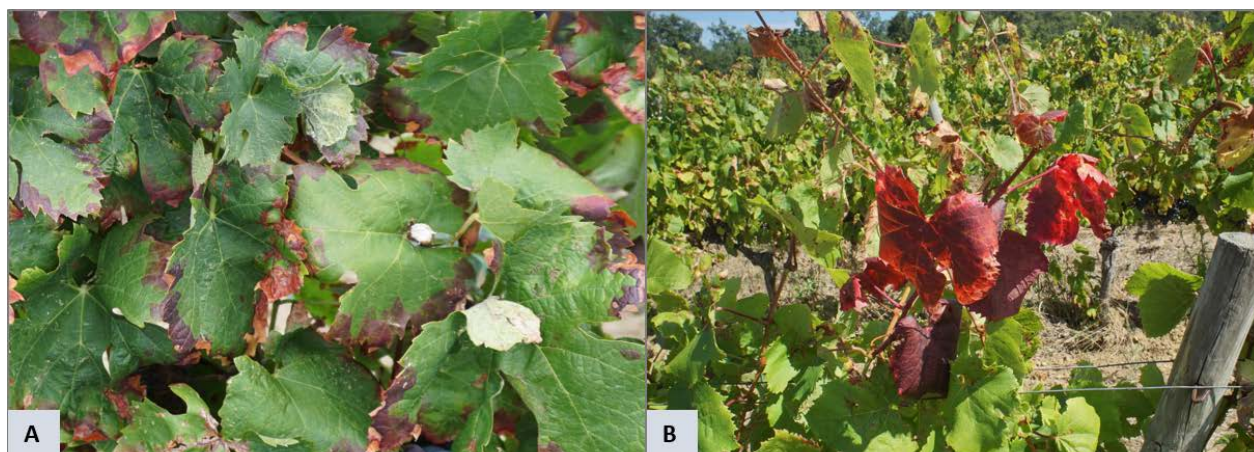


FIGURE 6.15 – Dégâts causés par des cicadelles : (A) Espèce non connue, (B) Cicadelle bubale

Ravageurs En plus des cicadelles dont les dégâts ont été montrés précédemment, deux autres types de ravageurs ont été observés : l'acarien (vers Carcassonne, Limoux et Cognac) et le scarabée (principalement en Montérégie). Les acariens causent notamment l'érinose (Figure 6.19, A) mais laissent aussi des taches jaunes et rouges (Figure 6.19, B). Le scarabée principalement observé est le scarabée japonais qui laisse des trouées dans les feuilles (Figure 6.19, C).

Résidus de traitements Les traitements réalisés par les viticulteurs, que ce soit l'apport d'engrais ou de pesticides, peuvent laisser des traces sur le feuillage. Dans le cas d'un traitement suivi par un fort ensoleillement, le feuillage peut être brûlé (Figure 6.20, A). Les traitements peuvent aussi laisser des résidus en quantité variable, souvent de couleur blanche ou bleutée (Figure 6.20, B et C).

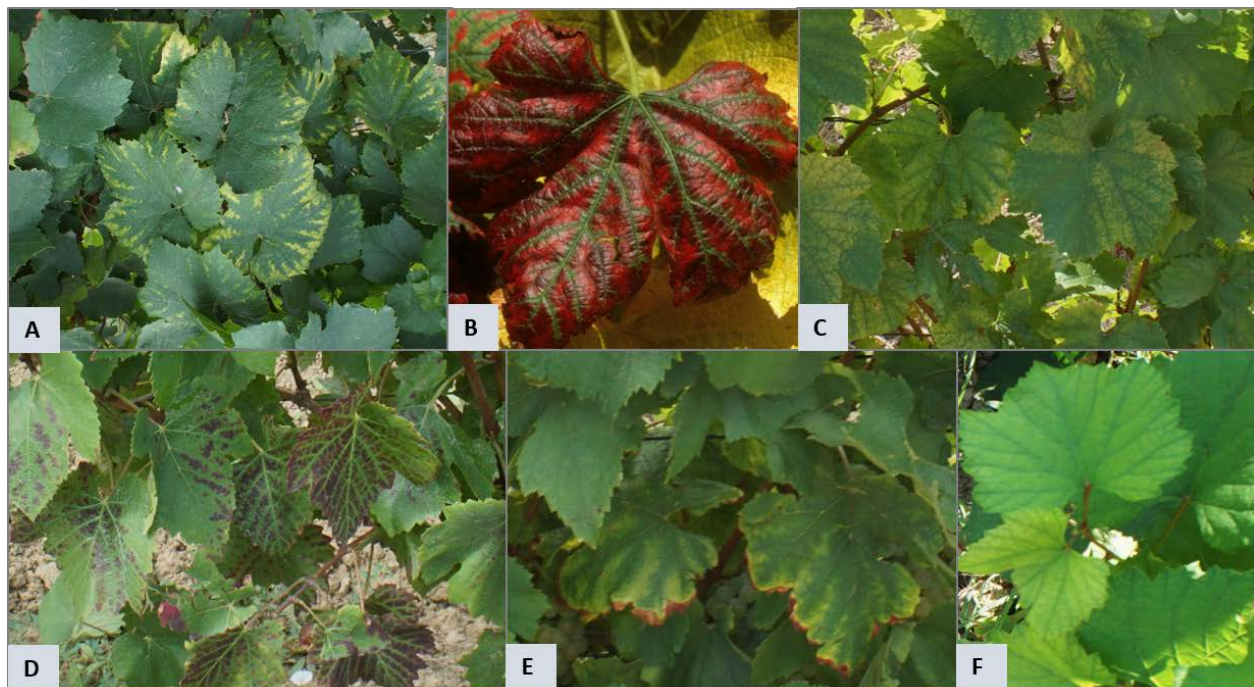


FIGURE 6.16 – Carences : (A) En magnésium sur cépage blanc. (B) En magnésium sur cépage rouge. (C) En manganèse. (D) En potassium sur cépage rouge. (E) En potassium sur cépage blanc. (F) En fer (peu avancé).



FIGURE 6.17 – Esca sur : (A) Cépage rouge. (B) Cépage rouge. (C) Cépage blanc.



FIGURE 6.18 – Black Dead Arm : (A) (B) (C) Trois expressions foliaires différentes. (D) Forme apoplectique de BDA ou d’esca.



FIGURE 6.19 – Dégâts provoqués par des ravageurs : (A) Erinose. (B) Marques laissées par des acariens. (C) Marques laissées par des scarabées japonais.



FIGURE 6.20 – Marques laissées par les traitements : (A) Phytotoxicité. (B) et (C) Embruns de traitements.

Problèmes phytosanitaires non identifiés L'identification de problèmes phytosanitaires est une entreprise complexe. Il n'a pas toujours été possible d'associer un problème aux symptômes observés, même en présence des inspecteurs de la Chambre d'Agriculture et de la FEDON. Cela pouvait notamment être dû à une expression particulière ou incomplète des symptômes. Les symptômes étaient également parfois trop peu développés pour qu'un diagnostic puisse être établi. En Figure 6.21, quelques exemples de cas non identifiés sont regroupés.

Conditions d'acquisition

Une volonté lors des acquisitions était d'intégrer le plus de diversité en termes de conditions d'illumination et d'angles de prise de vue dans l'objectif de rendre le modèle entraîné insensible à ces variations. Aucune restriction en termes de conditions météorologiques n'a été fixée. La seule exigence était que les feuilles soient sèches. Lors des acquisitions, à la fois les rangs à l'ombre et au soleil étaient observés, ce qui a permis de collecter des images aux expositions variées (Figure 6.22). L'angle de prise de vue ainsi que la proximité de la caméra et du feuillage varient d'une acquisition à l'autre (Figure 6.23). Ces paramètres dépendaient de la configuration du vignoble (taille de l'inter-rang) ainsi que de la localisation des organes d'intérêt sur le pied. La visée latérale a été adoptée de façon majoritaire.



FIGURE 6.21 – Exemples de problèmes phytosanitaires non identifiés.

Organisation des données

Les images sont organisées dans des dossiers assignés à la journée et la parcelle d'acquisition. Ce type de séparation permet d'assurer l'indépendance entre images lors de la constitution des ensembles d'entraînement / validation et de test. Les images ont ensuite été intégrées dans d'autres sous-dossiers correspondants au statut phytosanitaire des plantes de l'image (flavescence dorée / non flavescence dorée). Cette organisation par parcelle n'a néanmoins pas pu être appliquée pour les images prises à la main à Frelisghburg en 2017. En effet, ces images n'ont pas été acquises en vue de ce projet. Leur organisation avait été effectuée en fonction de la date d'acquisition et du statut phytosanitaire (mildiou / non mildiou). Il n'a pas été possible de procéder à une réorganisation par parcelle.



FIGURE 6.22 – Exemples de différentes conditions d’illumination rencontrées.



FIGURE 6.23 – Différents angles d’acquisition et distance au feuillage : (A) Visée nadir, acquisition éloignée. (B) Visée nadir, acquisition rapprochée. (C) Visée latérale, acquisition éloignée. (D) Visée latérale, acquisition rapprochée.

Annexe B : Détails techniques

Équipement informatique

Deux ordinateurs ont été utilisés pour l'entraînement et l'évaluation des modèles. Leurs caractéristiques sont regroupées dans la Table 6.3.

TABLEAU 6.3 – Caractéristiques des ordinateurs utilisés pour entraîner et évaluer les modèles.

	PC1	PC2
Processeur (CPU)	i5-7500	i9-7980XE
Mémoire vive (en Go)	32	128
Processeur graphique (GPU)	Nvidia, GeForce GTX 1050	Nvidia, GeForce GTX 1080
Nombre de GPU	1	3

Librairies et scripts

L'ensemble des scripts utilisés et développés sont en langage Python. Le *framework* ayant servi à entraîner et évaluer les modèles a été développé au CRIM. Il est disponible au lien suivant : <https://github.com/plstcharles/thelper>. Les principales librairies utilisées sont regroupées dans la Table 6.4.

TABLEAU 6.4 – Principales librairies utilisées dans le développement des algorithmes

Librairie	Tâche associée	Documentation
PyTorch	Apprentissage profond	https://pytorch.org
OpenCV	Pré-traitements et augmentations	https://pypi.org/project/opencv-python/
Augmentor	Augmentations	https://augmentor.readthedocs.io
Captum	Visualisation (Guided Grad-CAM)	https://captum.ai
Umap-learn	Visualisation (UMAP)	https://umap-learn.readthedocs.io
FastAI	Définition du taux d'apprentissage	https://fastai1.fast.ai

Temps d'entraînement et d'inférence

Selon la disponibilité du matériel, les entraînements et évaluations ont été réalisés sur l'un ou l'autre des ordinateurs présentés précédemment. Les temps d'entraînement des ResNet réalisés lors des travaux pour la reconnaissance du mildiou de la vigne sont rapportés dans la Table 6.5. Pour rappel, ces entraînements ont été réalisés sur un ensemble de 7 568 images pour 30 époques avec une taille de batch de 32 images.

Dans le cas d'étude pour la reconnaissance de symptômes de FD, l'entraînement de chaque ResNet-18 a duré environ 30mn sur le PC2. Pour rappel, l'entraînement s'effectuait sur un ensemble de 7 346

TABLEAU 6.5 – Temps d’entraînement des ResNet pour le cas d’étude sur la reconnaissance du mildiou

Architecture	id PC	Durée de l’entraînement
ResNet-18	1	1h01mn
ResNet-34	1	1h43mn
ResNet-50	2	1h
ResNet-101	2	1h11mn
ResNet-152	2	1h24mn

images, pour 30 époques, avec une taille de batch de 64 images. L’inférence en RPC durait environ 6s pour une image de 3 976 par 2 652 pixels.

Dans le cas d’étude évaluant les performances des RNCs face à un changement de résolution spatiale, ce sont également des ResNet-18 qui ont été entraînés. Les mêmes hyperparamètres que dans le cas d’étude de reconnaissance de symptôme de FD ont été utilisés. Les temps d’entraînement en fonction du jeu de données utilisé sont regroupés dans la Table 6.6. Quant aux durées d’inférence selon la dimension des images, elles sont reportées dans la Table 6.7.

TABLEAU 6.6 – Temps d’entraînement des ResNet-18 selon le jeux d’entraînement utilisé dans le cas d’étude évaluant les performances des RNCs face à un changement de résolution spatiale.

Ensemble de données	Nombre d’images	id PC	Durée de l’entraînement
Original – Ch4.	7 386	2	30mn
Approche 1	8 872	1	41mn
Approche 2	17 744	1	1h14mn

TABLEAU 6.7 – Temps d’inférence par image selon la dimension des images dans le cas d’étude évaluant les performances des RNCs face à un changement de résolution spatiale. Les inférences ont été réalisées sur le PC1.

Ensemble de données	Dimensions (pixels)	Durée d’inférence (s)
Drone 10m	3 976 * 2 652	4
Drone 25m	1 616 * 1 078	1
Perche	3 976 * 2 652	3
Perche intermédiaire	1 729 * 1 153	1
Perche éq. drone	857 * 571	1

Le rouge des fraises
Dans la tête, le vert des poires

Bleu fuschia, Odezenne

